2024년 경제활동인구조사
8차 연동 표본설계 보고서

2023. 10.

통계청 조사관리국
표 본 과

목 차

Ⅰ. 경제활동인구조사 개요 ····· 1

Ⅱ. 경제활동인구조사 표본설계 특징 ····· 2

Ⅲ. 경제활동인구조사 표본설계 ····· 4

1. 모집단 및 표본추출틀 ····· 4
2. 층화 및 분류지표 ····· 6
3. 표본 규모 ····· 11
4. 표본 추출 ····· 17

Ⅳ. 가중값 작성 및 추정 ····· 18

1. 가중값 작성 ····· 18
2. 추정 ····· 19

Ⅰ 경제활동인구조사 개요

1. 조사 목적

○ 국민의 경제활동(취업자, 실업자, 비경제활동인구 등) 규모와 특성을 파악하여 국가의 고용정책 수립과 평가에 필요한 자료 작성

- 노동공급, 고용구조, 가용 노동시간 및 인력자원의 활용 정도를 제공하고 일자리 창출과 실업자 지원 정책 등의 기초자료로 활용

○ 법적근거: 통계법 제 17조 1항에 의한 지정통계(승인번호 제101004호)

2. 조사 대상

○ 대한민국에 상주하고 있는 만 15세 이상 모든 사람

- 전국 1,791개 표본조사구에 거주하는 만 15세 이상 가구원(약 58,000명)
- 제외자: 군인(직업군인, 상근예비역 포함), 사회복무요원, 의무경찰 등

3. 조사 기간

○ 조사대상기간: 매월 15일이 포함된 1주 7일간(일~토요일)

○ 조사기간: 조사대상기간 다음 1주간

4. 조사 및 작성 주기

○ 조사: 매월 조사

- 부가 조사: 청년층 및 고령층(5월), 근로형태별(8월), 비임금근로자(8월), 비경제활동인구(8월) 실시

○ 공표: 월간 및 연간 공표

5. 조사 방법

○ 면접조사 및 전자조사(CASI)를 통한 자료 수집

Ⅱ 경제활동인구조사 표본설계 특징

1. 다목적 표본

○ 표본관리의 효율성, 행정 편의 등을 고려하여 1969년부터 다목적 표본제*를 도입하여 경제활동인구조사(이하 경활)와 집세조사 동시 실시

- 전국 17개 시도에서 경활을 실시하며, 소비자물가지수 작성 도시(40개)에서 집세조사 실시
- 동일한 표본 가구에서 조사하되, 조사기간을 달리하여 응답부담 경감

* 다목적 표본은 동일한 표본 조사구에서 여러 조사를 함께 실시하는 것을 말하며, 가계동향조사는 가계지출조사와 소득조사가 통합됨에 따라 2020년부터 제외

○ 경활 및 집세조사의 모집단 특성을 반영할 수 있는 표본설계 구현

- 경제활동 및 전·월세 특성과 연관성이 높은 변수와 거처유형을 분류지표로 활용하여 다목적 표본의 효율성을 높이고 모집단에 대한 대표성 확보
- 경활 표본 조사구 추출 후, 도시별 집세조사 최대 조사구역 설정 및 별도 조사구 추출

2. 연동 표본 체계

○ 표본 개편으로 인한 시계열 단절 방지를 위해 2005년부터 연동표본(rotation sample)을 도입하였으며, 2024년부터 8차 연동 표본 적용

- 매월 전체 표본의 1/36이 교체되며, 전월 대비 97.2%, 전년 대비 66.7% 중복

<표 Ⅱ-1> 연동표본 개편 연혁

	4차 연동	5차 연동	6차 연동	7차 연동	8차 연동
연동표본 교체시기	2013. 1.	2016. 1.	2018. 1.	2021. 1.	**2024. 1.**
표본 추출틀	2010년 인총 +신축조사구(11년)	2010년 인총 +신축조사구(14년)	2015년 등록센서스	2018년 등록센서스	**2021년 등록센서스**
표본규모 (조사구)	1,629개	1,647개 (경기, 제주 각 9개 확대)	1,737개 (세종 54개, 강원 18개, 전북 18개 확대)	1,737개	**1,791개** (세종, 경기, 충남, 전북, 경남, 제주 각 9개 확대)

3. 최신 모집단 반영

○ 2021년 등록센서스 자료와 20% 인총 표본조사 자료 활용

- 6차 연동부터 등록센서스 기반 표본추출틀 활용하여 최신 모집단 특성 반영
- 행정자료 기반 전수자료(주소, 거처, 가구·인구 DB)와 최신 인총 표본조사 자료를 주소(거처) 단위로 연계 및 공유하여 표본추출틀 구축

○ 모집단 변화를 적시 반영하기 위해 매년 표본 추출

- 등록센서스 기반 표본추출틀을 활용한 6차 연동부터 매년 표본조사구 추출
  · 1년 동안 전체 표본의 1/3이 교체되며, 2024년에는 597개 조사구 진입 예정(8차 연동 4~6그룹)
- 연동표본 개편 이후 3년 동안은 표본설계 및 추출 방법 유지
  · 층화 기준, 표본규모, 설계층별 분류지표(정렬변수)를 동일하게 적용

4. 고용특화 표본추출틀 구축 및 활용

○ 취업활동 통계등록부를 활용한 표본추출틀 구축(7차 연동부터 활용)

- 사회보험, 과세자료 및 가족관계등록부 등을 이용하여 구축한 2021년 기준 취업활동 통계등록부(개인 단위)를 2021년 등록센서스 기반 조사구와 연계
- 조사구별 고용률, 산업별 취업자 비율 등 고용 특성을 나타내는 변수 생성

○ 등록 고용률 및 취업자의 산업분포를 고려한 연동 그룹 결정 및 대체 조사구 선정을 통해 안정적인 표본교체 도모

III 경제활동인구조사 표본설계

1. 모집단 및 표본추출틀

가. 모집단 정의

○ **목표 모집단**: 대한민국에 상주하고 있는 만 15세 이상 모든 사람

- 만 15세 이상 인구 중 아파트 조사구에 거주하는 인구가 54.2%로 가장 많고, 보통 조사구 42.0%, 섬 조사구 0.2%로 나타났으며, 기숙/사회시설에는 1.9%가 거주하는 것으로 나타남('21년 등록센서스 기준)

○ **조사 모집단**: '21년 기준 등록센서스에 포함된 만 15세 이상 모든 사람

- 병영(군대), 병원, 교도소 등 특수시설 거주자와 불법체류자는 조사모집단에서 제외하며, 등록센서스 조사구 정보와 연계되지 않는 경우도 제외
- 목표 모집단의 1.7%(778,577명) 제외

<표 Ⅲ-1> 모집단 현황

(단위: 개, 명, %)

조사구 특성	2020년 기준			2021년 기준		
	조사구	만 15세 이상 인구	구성비	조사구	만 15세 이상 인구	구성비
전체	384,635	45,574,979	100.0	389,814	45,650,600	100.0
아파트(A)	195,840	24,402,401	**53.5**	199,746	24,764,357	**54.2**
보통(1)	187,882	19,355,932	**42.5**	189,156	19,178,860	**42.0**
섬(2)	913	69,339	**0.2**	912	67,413	**0.2**
기숙/사회시설[1)]	-	871,929	**1.9**	-	861,393	**1.9**
기타(제외)[2)]	-	875,378	1.9	-	778,577	1.7

1) 거처유형이 '기숙사' 또는 '사회시설'이면서 가구구분이 '집단시설가구'인 경우(조사구 번호 없음)
2) 군인, 해외공관, 불법체류자 등 제외자 및 등록센서스 조사구 정보와 연계되지 않는 경우

나. 표본추출틀 구축

○ 표본추출틀 현황

- ’21년 기준 등록센서스 조사구 중에서 아파트(A), 보통(1) 및 섬(2) 조사구를 대상으로 구축

- 조사모집단 내 비중이 낮고 현장 조사가 어려운 기숙/사회시설과 인구 및 가구가 없는 빈집 조사구 304개 제외, 총 389,510개 조사구

○ **포함률**: 조사구 기준 99.9%, 15세 이상 인구 기준 98.1%

- 시도별 포함률도 96.6~99.0%로 조사모집단에 대한 대표성이 높음

<표 Ⅲ-2> 조사모집단 대비 표본추출틀 현황

(단위: 개, 명, %)

지역	조사모집단		제외		표본추출틀			
	조사구(A)	15세 이상 인구(B)	빈집 조사구	기숙/시설 인구*	조사구(a)	포함률(a/A)	15세 이상 인구(b)	포함률(b/B)
전국	**389,814**	44,872,023	**304**	861,393	**389,510**	***99.9***	**44,010,630**	***98.1***
서울	72,343	8,416,931	9	82,565	72,334	100.0	8,334,366	99.0
부산	26,550	2,935,647	6	61,441	26,544	100.0	2,874,206	97.9
대구**	17,385	2,103,686	5	36,313	17,380	100.0	2,067,373	98.3
인천	21,000	2,557,104	10	34,277	20,990	100.0	2,522,827	98.7
광주	10,598	1,265,354	2	27,965	10,596	100.0	1,237,389	97.8
대전	10,956	1,284,169	65	35,365	10,891	99.4	1,248,804	97.3
울산	8,339	957,032	26	17,915	8,313	99.7	939,117	98.1
세종	2,667	290,440	2	5,597	2,665	99.9	284,843	98.1
경기	93,950	11,661,418	76	175,275	93,874	99.9	11,486,143	98.5
강원	12,957	1,332,156	16	34,195	12,941	99.9	1,297,961	97.4
충북	12,839	1,405,301	2	40,082	12,837	100.0	1,365,219	97.2
충남	17,007	1,867,878	7	56,532	17,000	100.0	1,811,346	97.0
전북	14,407	1,560,555	11	48,721	14,396	99.9	1,511,834	96.9
전남	15,241	1,547,443	6	47,580	15,235	100.0	1,499,863	96.9
경북**	21,950	2,282,427	26	78,019	21,924	99.9	2,204,408	96.6
경남	26,212	2,839,453	34	69,195	26,178	99.9	2,770,258	97.6
제주	5,413	565,029	1	10,356	5,412	100.0	554,673	98.2

* 기숙/시설 인구: 거처유형이 ‘기숙사 및 사회시설’이고 가구구분이 ‘집단시설가구’인 가구에 거주하는 만 15세 이상 인구
** 2023. 7. 1. 대구광역시로 편입된 경북 군위군의 214개 조사구는 대구광역시로 집계

2. 층화 및 분류지표

가. 층화

○ 27개 지역층 검토

- 공표 수준과 시계열 안정성을 고려하여 기존 27개 설계층 검토

· 공표 수준: 전국, 17개 시도

· 7차 연동 설계층: 7개 특·광역시, 세종 및 9개 도 지역(동부, 읍면부)

○ 인구 사회학적 분포를 고려하여 27개 설계층 유지

- '21~'22년 경제활동인구조사의 경활상태와 연관성이 높은 인구·사회학적 특성 변수의 분포가 설계층별로 큰 차이가 나타나 기존 설계층 유지

* 설계층별 인구·가구 특성 분포는 미공표 통계이므로 첨부하지 않음

<표 Ⅲ-3> 시도별 층 수 및 층화 기준

시도	층 수	재층화
서울, 부산, 대구, 인천, 광주, 대전, 울산	**7개**	재층화 없음
세종, 경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주	**20개**	동/읍면부로 재층화

나. 분류 지표

○ 조사항목과 연관성이 높은 특성 변수(분류지표)에 따라 조사구를 정렬한 후 계통추출하면 **내재적 층화효과(implicit stratification)**를 얻을 수 있으므로, 경제활동 상태와 연관성이 높은 변수 선정 필요

1) 분석 방법

○ 자료 연계

- 모집단의 인구·가구·주거 특성과 경제활동 상태의 연관성 분석을 위해 '21년 조사자료를 '21년 기준 등록센서스 자료와 연계

· '21년 표본은 연동모형에 의해 '15년~'19년 기준 표본추출틀에서 추출된 표본으로 구성되어 있으므로 주소 정비 후 자료 연계, 연계율 98.48%

○ **분석 절차**

- 경제활동 참가 여부, 취업 여부와 조사구 특성 변수의 연관성 분석
  · 로지스틱 회귀분석을 통해 두 종속변수에 모두 유의미한 변수 선정
  · 상관관계가 높은 변수군 중 등록 고용률과 연관성이 높은 대표 변수 선별
- 다중 로지스틱 회귀분석을 통해 설계층별 상위 변수 3개 선정(연속형)
- 중앙값 기준으로 각각 2개 수준으로 범주화하여 분류지표 생성

○ **조사구 특성 변수**

- **지역 구분**
  · **(사무소)** 자치 시군구를 구분하여 관할
  · **(군집)** 세종을 제외한 16개 시도 내 시군구를 경활상태와 연관성이 높은 조사구 특성변수를 이용하여 2개 군집으로 분류
  · **(도시화 정도)** 인구밀도를 기준으로 도시/준도시/비도시로 분류
  · **(행정구역 지위)** 동/읍/면
- **고용, 인구·가구 특성 변수**
  · **(고용)** 등록 고용률, 농림어업 종사자 비율(인총, 등록), 광제조업 종사자 비율(인총, 등록), 서비스업 종사자 비율(인총, 등록)
  · **(인구)** 평균 연령, 15세 이상 인구 비율, 15~29세 인구 비율, 30~49세 인구 비율, 65세 이상 인구 비율, 70세 이상 인구 비율, 대졸 이상 비율, 3~40대 인구 중 대졸 이상 비율, 30세 이상 인구 중 대졸 이상 비율, 유배우 인구 비율, 노년 부양비
  · **(가구)** 가구주 평균 연령, 25~49세 가구원 포함 가구 비율, 1인 가구 비율, 65세 이상 인구만 있는 가구 비율, 농림어가 비율, 외국인 가구 비율
- **주거 특성 변수**
  · **(주거 점유)** 자가 소유 가구 비율(행정자료), 자가 가구 비율(인총), 전세 가구 비율(인총), 월세 가구 비율(인총)
  · **(거처 유형)** 27개 설계층 내에서 3~5개 범주로 구분

2) 1~2차 분류지표: 지역 구분

○ 특정 시군구에서 표본이 과다 또는 과소 추출되는 것을 방지하기 위하여, 인구·사회학적 특성이 유사한 지역을 묶어 분류지표로 활용

- 설계층별로 사무소, 군집, 도시화 정도 중 설명력이 높은 변수 선정(최대 2개)
  · 도시화 정도는 조사구 빈도를 고려하여 일부 지역은 2개 그룹으로 병합
  · 행정구역 지위(동/읍/면)보다 도시화 정도가 통계적으로 유의미한 설계층이 많아 동/읍/면은 제외

<표 Ⅲ-4> 1~2차 분류지표 - 지역 구분

설계층		분류지표 순서	군집	도시화 정도
서울		사무소	-	-
부산		군집 > 도시화 정도	군집1: 강서구, 기장군 군집2: 강서구, 기장군 외 14개 구	도시/준도시·비도시(2)
대구		군집 > 도시화 정도	군집1: 군위군 군집2: 군위군 외 8개 구·군	도시/준도시·비도시(2)
인천		군집 > 도시화 정도	군집1: 동구, 강화군, 옹진군 군집2: 동구, 강화군, 옹진군 외 7개 구	도시/준도시·비도시(2)
광주		군집 > 도시화 정도	군집1: 광산구 군집2: 광산구 외 4개 구	도시/준도시·비도시(2)
대전		군집 > 도시화 정도	군집1: 서구, 유성구 군집2: 서구, 유성구 외 3개 구	도시/준도시·비도시(2)
울산		군집 > 도시화 정도	군집1: 북구 군집2: 북구 외 4개 구·군	도시/준도시/비도시(3)
세종	동	지역 구분×	-	-
	읍면	도시화 정도	-	도시/준도시/비도시(3)
경기	동	사무소 > 군집	군집1: 수원시 권선구, 영통구, 안산시, 오산시, 시흥시, 하남시 군집2: 그 외 21개 행정구	-
	읍면	사무소 > 도시화 정도	-	도시/준도시/비도시(3)
강원	동	사무소 > 도시화 정도	-	도시/준도시·비도시(2)
	읍면	사무소 > 도시화 정도	-	도시·준도시/비도시(2)
충북	동	사무소 > 군집	군집1: 충주시, 제천시 군집2: 청주시 4개 구	-
	읍면	사무소 > 도시화 정도	-	도시·준도시/비도시(2)
충남	동	사무소 > 군집	군집1: 천안시, 당진시 군집2: 그 외 6개 시	-
	읍면	사무소 > 도시화 정도	-	도시/준도시/비도시(3)
전북	동	사무소 > 도시화 정도	-	도시/준도시·비도시(2)
	읍면	사무소 > 도시화 정도	-	도시·준도시/비도시(2)
전남	동	사무소 > 도시화 정도	-	도시/준도시·비도시(2)
	읍면	사무소 > 도시화 정도	-	도시/준도시/비도시(3)

<표 Ⅲ-4> 1~2차 분류지표 - 지역 구분(계속)

설계층		분류지표 순서	군집	도시화 정도
경북	동	사무소 > 도시화 정도	-	도시/준도시·비도시(2)
	읍면	사무소 > 도시화 정도	-	도시/준도시/비도시(3)
경남	동	사무소 > 군집	군집1: 진주시, 통영시, 사천시, 밀양시, 창원시 마산합포구, 마산회원구 군집2: 그 외 6개 행정구	-
	읍면	사무소 > 도시화 정도	-	도시/준도시/비도시(3)
제주	동	도시화 정도 > 군집	제주시, 서귀포시	도시/준도시·비도시(2)
	읍면	군집	제주시, 서귀포시	-

3) 3차 분류지표: 거처유형

○ 경활 표본 조사구에서 함께 조사하는 집세조사를 고려하여 거처유형을 분류지표로 활용

- 경남 읍면부와 제주 읍면부를 제외한 25개 설계층에서 거처유형과 경활상태의 연관성이 높은 것으로 나타남

<표 Ⅲ-5> 3차 분류지표 - 거처유형

거처유형		설계층			
		7개 특·광역시, 경기/충남/전북 동부	세종 동부	강원/충북/전남/경북/경남/제주 동부 세종/경기/강원/충북/충남/전북/전남/경북/경남 읍면부	제주 읍면부
단독	일반, 다가구, 영업겸용	① 단독주택	① 단독/기타	① 단독주택	① 단독주택
기타	연립, 다세대, 기타 주택	⑤ 기타		④ 기타	③ 기타
	주택 외 거처				
아파트	전용면적 60㎡ 이하	② 소형APT	② 소형APT	② 소형APT	② APT
	60㎡ 초과 ~ 102㎡ 이하	③ 중형APT	③ 중형APT	③ 중대형APT	
	102㎡ 초과	④ 대형APT	④ 대형APT		

4) 4~6차 분류지표: 인구 · 가구 특성

○ 고용, 인구·가구, 주거 특성 반영

- 경제활동 참가 여부와 취업 여부 모두 유의미한 변수 중 설명력이 높은 변수를 3개 선정하여 분류지표 생성
- 다중 로지스틱 회귀분석을 통해 변수를 선정하고, 중앙값을 기준으로 범주화

<표 Ⅲ-6> 4~6차 분류지표

설계층		인구·가구 특성 변수
서울		등록 고용률 > 가구주 평균연령 > 광제조업 종사자 비율(인총)
부산		70세 이상 인구비율 > 등록 고용률 > 전세가구 비율(인총)
대구		등록 고용률 > 농림어가 비율 > 광제조업 종사자 비율(인총)
인천		등록 고용률 > 25~49세 가구원 포함비율 > 농림어가 비율
광주		등록 고용률 > 자가가구 비율(인총) > 서비스업 종사자 비율(인총)
대전		등록 고용률 > 서비스업 종사자 비율(인총) > 25~49세 가구원 포함비율
울산		등록 고용률 > 대졸이상 비율_전체인구중 > 65세이상 인구비율
세종	동	등록 고용률 > 30~49세 인구비율 > 70세이상 인구비율
	읍면	등록 고용률 > 25~49세 가구원 포함비율 > 가구주평균연령
경기	동	가구주 평균연령 > 등록 고용률 > 대졸이상 비율_전체인구중
	읍면	자가가구 비율(인총) > 가구주 평균연령 > 농림어가 비율
강원	동	등록 고용률 > 자가가구 비율(인총) > 전세가구 비율(인총)
	읍면	등록 고용률 > 전세가구 비율(인총) > 25~49세 가구원 포함비율
충북	동	70세이상 인구비율 > 전세가구 비율(인총) > 등록 고용률
	읍면	25~49세 가구원 포함비율 > 광제조업 종사자 비율(인총) > 30~49세 인구비율
충남	동	등록 고용률 > 광제조업 종사자 비율(인총) > 전세가구 비율(인총)
	읍면	등록 고용률 > 70세이상 인구비율 > 가구주 평균연령
전북	동	등록 고용률 > 65세이상 인구비율 > 농림어가 비율
	읍면	자가가구 비율(인총) > 대졸이상 비율_전체인구중 > 가구주 평균연령
전남	동	등록 고용률 > 15~29세 인구비율 > 70세이상 인구비율
	읍면	30~49세 인구비율 > 65세이상 인구비율 > 등록 고용률
경북	동	등록 고용률 > 70세이상 인구비율 > 15~29세 인구비율
	읍면	농림어가 비율 > 대졸이상 비율_전체인구중 > 25~49세 가구원 포함비율
경남	동	등록 고용률 > 가구주 평균연령 > 광제조업 종사자 비율(인총)
	읍면	농림어가비율 > 70세이상 인구비율 > 자가가구 비율(인총)
제주	동	등록 고용률 > 15~29세 인구비율 > 70세이상 인구비율
	읍면	서비스업 종사자 비율(인총) > 광제조업 종사자 비율(인총) > 15세이상 인구비율

3. 표본 규모

가. 표본 가구 수

○ 조사구당 최적 표본 가구 수

- 현행 경제활동인구조사의 조사구 당 표본 가구 수(20가구)의 적절성 검토

$$m_{opt} = \sqrt{\frac{S_w^2}{S_b^2}\frac{C_{psu}}{C_{ssu}}}$$

· $S_w^2 = MSW = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\overline{y_i})^2$: 집락 내 변동

· $S_b^2 = \frac{MSB-MSW}{m}$: 집락 간 변동

· $MSB = \frac{m}{n-1}\sum_{i=1}^{n}(\overline{y_i}-\overline{\overline{y}})^2$

· m_{opt} : 최적 집락크기(가구 수)

· C_{psu} : 조사구당 조사비용

· C_{ssu} : 조사구 내 가구당 조사비용

· n : 표본 조사구 수

· m : 표본 조사구 내 가구 수

- 조사구당 조사비용(C_{psu})과 조사구 내 가구당 조사비용(C_{ssu})의 상대적인 비율($C_{psu} : C_{ssu}$)을 가정한 후, 조사구 내 최적 가구 수 산출

· C_{psu} : 조사구 조사비용 - 연동교체 조사담당자 여비 등

· C_{ssu} : 가구 조사비용 - 답례품, 면접비용, 조사표류 인쇄비 등

· 조사비용 비율: '22년 기준 8:1 ~ 10:1 수준으로 추정

- '22년 조사자료를 이용하여 주요 변수별 최적 가구 수 산출

· 조사구에서 가구를 추출하므로 가구 기준 경활인구, 취업인구, 비경활 인구, 실업인구 변수 생성

· 설계층 및 월별 주요 변수의 S_w^2, S_b^2를 추정하고 최적 가구 수의 중앙값 요약

<표 Ⅲ-7> 설계층별 조사구당 최적 표본 가구 수

(단위: 가구)

$C_{psu}:C_{ssu}$		8:1				10:1				12:1			
특성		경활	취업	실업	비경활	경활	취업	실업	비경활	경활	취업	실업	비경활
평균		16.6				18.5				20.3			
		13.6	13.5	25.3	13.9	15.2	15.1	28.3	15.6	16.6	16.5	31.0	17.1
서울		10.7	10.7	22.6	9.8	12.0	12.0	25.2	11.0	13.1	13.1	27.7	12.0
부산		15.2	14.1	28.2	19.8	17.0	15.8	31.5	22.1	18.6	17.3	34.5	24.2
대구		9.9	9.7	32.6	11.4	11.1	10.8	36.5	12.7	12.2	11.9	39.9	13.9
인천		9.1	9.2	27.4	11.9	10.2	10.3	30.6	13.4	11.2	11.2	33.5	14.6
광주		9.3	9.1	28.6	12.9	10.4	10.2	32.0	14.5	11.4	11.2	35.1	15.8
대전		9.3	9.0	25.5	13.2	10.4	10.0	28.6	14.8	11.3	11.0	31.3	16.2
울산		10.7	10.5	31.7	15.7	12.0	11.7	35.5	17.6	13.1	12.8	38.9	19.2
세종	동	28.1	29.4	26.0	10.8	31.4	32.9	29.0	12.1	34.4	36.0	31.8	13.3
	읍면	20.1	19.7	28.5	7.4	22.5	22.0	31.9	8.3	24.6	24.1	35.0	9.1
경기	동	12.5	12.3	31.8	9.1	14.0	13.7	35.5	10.2	15.3	15.0	38.9	11.1
	읍면	9.6	9.6	21.4	10.7	10.7	10.7	23.9	12.0	11.8	11.7	26.2	13.1
강원	동	14.5	13.9	20.9	11.2	16.2	15.5	23.3	12.5	17.7	17.0	25.6	13.7
	읍면	17.9	17.8	18.1	14.9	20.0	19.9	20.2	16.7	21.9	21.8	22.1	18.3
충북	동	10.9	10.7	22.2	9.7	12.2	12.0	24.8	10.9	13.4	13.1	27.2	11.9
	읍면	16.3	15.7	36.3	17.3	18.2	17.6	40.6	19.3	20.0	19.3	44.5	21.1
충남	동	10.2	10.0	20.1	9.5	11.4	11.2	22.4	10.7	12.5	12.2	24.6	11.7
	읍면	10.4	10.7	27.0	11.0	11.7	12.0	30.2	12.3	12.8	13.1	33.1	13.4
전북	동	11.8	11.3	21.2	20.7	13.2	12.7	23.7	23.2	14.5	13.9	25.9	25.4
	읍면	22.3	23.0	23.4	19.6	25.0	25.8	26.2	21.9	27.3	28.2	28.7	24.0
전남	동	12.6	13.5	26.1	25.8	14.1	15.1	29.2	28.8	15.4	16.5	32.0	31.5
	읍면	16.4	16.9	30.9	16.0	18.3	18.9	34.5	17.9	20.1	20.7	37.8	19.7
경북	동	9.3	9.5	18.0	12.6	10.4	10.6	20.1	14.0	11.4	11.6	22.0	15.4
	읍면	12.3	11.9	23.3	14.6	13.7	13.3	26.0	16.3	15.0	14.6	28.5	17.9
경남	동	12.0	12.2	21.2	12.7	13.5	13.7	23.8	14.1	14.7	15.0	26.0	15.5
	읍면	14.7	13.1	21.5	15.2	16.5	14.7	24.0	17.0	18.1	16.1	26.3	18.7
제주	동	10.0	10.3	23.3	13.5	11.2	11.6	26.0	15.1	12.2	12.7	28.5	16.6
	읍면	19.9	20.0	25.5	19.3	22.2	22.4	28.5	21.6	24.4	24.5	31.3	23.7

☞ 적정 표본 가구 수는 지역 및 변수별 차이가 있으나,
조사비용 비율별 **평균이 16.6~20.8가구 수준**이고
조사담당자 업무량과 연동표본체계 유지를 위해 **20가구로 결정**

나. 표본 조사구 수

○ 시도별 표본 배분방법 검토

- 전국 1,791개 조사구를 15세 이상 인구 기준으로, 17개 시도별 비례배분 및 제곱근 비례배분 방법 검토

· 제곱근 비례배분 결과, 부산, 대구, 인천, 경기, 경남 지역의 표본은 확대하고, 세종, 강원, 전북, 제주 표본 감소 필요(표 Ⅲ-9)

<표 Ⅲ-8> 7차 연동 표본규모 및 시도별 상대표준오차

(단위: 개, %)

설계층		7차 연동 표본 조사구 수			'22년 연간 RSE		'22년 월별 RSE 평균	
		'21~'22년	'23년	확대	취업자	실업자	취업자	실업자
전국		1,737	1,791	54	0.3	2.6	0.4	4.2
서울		207	207		0.7	8.0	1.0	10.1
부산		108	108		0.9	5.8	1.2	14.8
대구		90	90		1.0	6.0	1.5	15.9
인천		99	99		1.0	6.1	1.2	13.0
광주		81	81		1.0	8.4	1.8	14.9
대전		81	81		1.1	14.9	1.5	19.4
울산		72	72		1.2	7.2	1.5	16.3
세종	동	27	36	9	1.9	12.1	2.1	23.6
	읍면	27	27					
경기	동	171	180	9	0.7	5.5	0.9	10.2
	읍면	36	36					
강원	동	63	63		1.1	8.0	1.5	15.8
	읍면	45	45					
충북	동	54	54		1.1	8.7	1.5	17.6
	읍면	36	36					
충남	동	45	45		1.2	8.2	1.7	18.0
	읍면	45	54	9				
전북	동	63	72	9	1.0	5.6	1.5	20.2
	읍면	36	36					
전남	동	36	36		0.8	4.6	1.4	18.2
	읍면	54	54					
경북	동	54	54		1.1	10.3	1.5	17.5
	읍면	54	54					
경남	동	63	63		1.1	6.5	1.8	15.3
	읍면	36	45	9				
제주	동	36	36		1.2	9.2	1.6	22.7
	읍면	18	27	9				

· 부산, 대구, 인천, 경기, 경남은 표본규모를 확대해도 RSE가 크게 감소하지 않고

· 세종, 전북, 제주는 '23년에 확대한 지역이므로 조정하지 않고 현재의 시도별 표본규모 유지

<표 Ⅲ-9> 시도별 표본규모 검토

(단위: 개, %)

설계층		7차 연동 표본 조사구 수		표본 조사구 수			제곱근 비례배분 연간 예상 RSE	
		'23년(a)	확대	비례배분	제곱근 비례배분(b)	차이 (b-a)	취업자	실업자
전국		1,791	54	1,791	1,791	0	0.3	2.5
서울		207		336	208	1	0.7	8.0
부산		108		117	123	15	0.9	5.4
대구		90		83	103	13	0.9	5.6
인천		99		102	115	16	0.9	5.6
광주		81		51	81	0	1.0	8.4
대전		81		51	81	0	1.0	14.9
울산		72		38	70	-2	1.2	7.3
세종	동	63	9	12	39	-24	2.2	14.3
	읍면							
경기	동	216	9	465	245	29	0.7	5.0
	읍면							
강원	동	108		53	83	-25	1.3	9.1
	읍면							
충북	동	90		56	85	-5	1.1	9.0
	읍면							
충남	동	99		75	98	-1	1.2	7.8
	읍면		9					
전북	동	108	9	62	89	-19	1.0	5.9
	읍면							
전남	동	90		62	89	-1	0.8	4.6
	읍면							
경북	동	108		92	109	1	1.1	10.2
	읍면							
경남	동	108		113	121	13	1.0	5.9
	읍면		9					
제주	동	63		23	54	-9	1.2	9.2
	읍면		9					

○ **세종 및 9개 도 지역의 표본 배분방법 검토**

- 15세 이상 인구 기준, 시도 내 동부·읍면부 표본의 비례배분 검토

· 비례배분 결과, 세종, 전북, 경남, 제주의 조정이 필요한 것으로 나타났으나, ’23년에 표본을 확대하였으므로, ’23년 추정결과를 검토한 후 결정 필요

<표 Ⅲ-10> 부차층별 표본규모 검토

(단위: 천 명, 개)

설계층		15세 이상 인구	시도 내 구성비(%)	시도별 표본 조사구수	동/읍면부 비례 배분	차이
세종	동	202	69.7	63	44	8
	읍면	88	30.3		19	-8
경기	동	9,678	83.0	216	179	-1
	읍면	1,983	17.0		37	1
강원	동	799	59.9	108	65	2
	읍면	534	40.1		43	-2
충북	동	811	57.7	90	52	-2
	읍면	595	42.3		38	2
충남	동	859	46.0	99	46	1
	읍면	1,009	54.0		53	-1
전북	동	1,108	71.0	108	77	5
	읍면	453	29.0		31	-5
전남	동	664	42.9	90	39	3
	읍면	884	57.1		51	-3
경북	동	1,223	53.1	108	57	3
	읍면	1,080	46.9		51	-3
경남	동	1,815	63.9	180	69	6
	읍면	1,024	36.1		39	-6
제주	동	404	71.4	63	45	9
	읍면	162	28.6		18	-9

<표 Ⅲ-11> 8차 연동 표본규모 및 시도별 예상 상대표준오차

(단위: 개, %)

설계층		7차 연동('23년) 표본 조사구 수		8차 연동('24년~) 표본 조사구 수				연간 예상 RSE	
		시도별	설계층별	시도별	설계층별	차이	그룹별	취업자	실업자
전국		**1,791**	**1,791**	**1,791**	**1,791**	**-**	**199**	**0.3**	**2.5**
서울		207	207	207	207	-	23	0.7	8.0
부산		108	108	108	108	-	12	0.9	5.8
대구		90	90	90	90	-	10	1.0	6.0
인천		99	99	99	99	-	11	1.0	6.1
광주		81	81	81	81	-	9	1.0	8.4
대전		81	81	81	81	-	9	1.1	14.9
울산		72	72	72	72	-	8	1.2	7.2
세종	동	63	36	63	36	-	4	1.7	11.2
	읍면		27		27	-	3		
경기	동	216	180	216	180	-	20	0.7	5.4
	읍면		36		36	-	4		
강원	동	108	63	108	63	-	7	1.1	8.0
	읍면		45		45	-	5		
충북	동	90	54	90	54	-	6	1.1	8.7
	읍면		36		36	-	4		
충남	동	99	45	99	45	-	5	1.2	7.8
	읍면		54		54	-	6		
전북	동	108	72	108	72	-	8	0.9	5.4
	읍면		36		36	-	4		
전남	동	90	36	90	36	-	4	0.8	4.6
	읍면		54		54	-	6		
경북	동	108	54	108	54	-	6	1.1	10.3
	읍면		54		54	-	6		
경남	동	108	63	108	63	-	7	1.1	6.2
	읍면		45		45	-	5		
제주	동	63	36	63	36	-	4	1.1	8.5
	읍면		27		27	-	3		

☞ 15세 인구 분포를 고려하여 표본 조사구 배분방법을 검토하였으나, '23년에 확대한 지역이 많아 현행 **설계층별 표본규모 유지**
'23~24년 추정결과를 검토한 후에 조정여부 결정 필요

4. 표본 추출

가. 표본추출방법

○ 층화 2단 집락 계통추출(Two-stage Cluster Systematic Sampling)

- 1차 추출단위: 조사구, 확률비례계통추출(가구 수 크기 비례)
- 2차 추출단위: 가구, 표본 조사구 내 단순임의추출

○ 모집단의 변화를 반영하기 위하여 전체 표본의 1/3 추출(597개 조사구)

- 매년 3개 그룹의 표본 조사구가 교체되며, '24년에는 8차 연동 4~6그룹 진입

○ 3개 그룹의 표본이 동질적인 분포를 갖도록 배분

- 설계층 내에서 지역, 거처유형, 등록 고용률과 분류지표의 분포가 그룹별로 유사한 분포를 갖도록 재추출하여 그룹 배정(Cube Method* 이용)

* Cube method(Deville & Tille, 2004)는 보조변수들에 대한 HT 추정량이 모집단과 일치하도록 근사적인 균형표본을 찾는 방법으로 최대 40개의 보조변수 활용 가능

나. 표본대체

○ 표본 조사구 유고 발생 시 조사 가능 조사구로 대체

- 조사구 특성이 유사한 표본 간 대체가 이루어질 수 있도록 분류지표 기준으로 적정 가구 수의 대체조사구 탐색

○ 조사구 대체 시 취업활동 통계등록부의 산업별* 취업자 비율을 고려하여 표본 대체로 인한 산업분포 변동 최소화

- 연동 교체 시 시도 단위 산업분포 변화가 크게 발생하는 문제 예방

* 농림어업, 광제조업, 건설업, 도소매음식숙박업, 전기수도운수업 및 기타서비스업, 무직으로 분류(6개 범주)

IV 가중값 작성 및 추정

1. 가중값 작성

가. 설계 가중값

○ 표본추출방법에 의해 설계층 내의 모든 표본의 설계 가중값 동일(자체가중)

- 조사구: 27개 설계층별 확률비례계통추출(가구 수 크기비례)
- 가구: 조사구 내에서 20가구 단순임의추출
- 가구원: 표본 가구의 만 15세 이상 가구원 전체

$$w_{hijk} = \underbrace{\frac{M_h}{n_h M_{hi}}}_{\substack{\uparrow \\ \text{조사구} \\ \text{포함확률} \\ \text{역수}}} \times \underbrace{\frac{M_{hi}}{m_{hi}}}_{\substack{\uparrow \\ \text{가구} \\ \text{포함확률} \\ \text{역수}}} = \frac{M_h}{n_h 20} \ (= w_h)$$

· w_{hijk} : h층 i조사구 j가구 k가구원의 설계 가중값
· M_h : h층의 총 가구 수
· M_{hi} : h층 i조사구의 가구 수(크기 척도)
· n_h : h층의 표본 조사구 수
· m_{hi} : 조사구 내 표본 가구 수($m_{hi} = 20$)

나. 사후 가중값

○ 표본설계에 반영하지 못한 인구 특성과 최신 모집단의 분포에 맞게 보정

- 27개 설계층의 성·연령 그룹별 현재인구 벤치마크, 총 702개 사후층

$$w_{hijk}^{ps} = w_{hijk} \times \frac{X_{h,p}}{\hat{X}_{h,p}}$$

· w_{hijk}^{ps} : h층 i조사구 j가구 k가구원의 사후 가중값
· X : 현재인구(벤치마크 모수)
· $\hat{X}$: 응답한 표본의 설계 가중값(w_{hijk})의 합
· p : 사후층, 성 및 연령 그룹(5세 단위)

2. 추정

가. 월간 추정

○ 총계: 사후 가중값을 이용한 사후층화 추정량으로 추정

$$\hat{t}_{y,post} = \sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{20}\sum_{k=1}^{u_j} w_{hijk}^{ps} \times y_{hijk}$$

· w_{hijk}^{ps} : h층 i조사구 j가구 k가구원의 사후 가중값

· y_{hijk} : h층 i조사구 j가구 k가구원의 관측값

- 사후층화 추정량은 <u>가구원 k의 사후층 포함 여부를 나타내는 지시변수들의 벡터를 보조 정보(x)로 사용하는 회귀 추정량으로 표현</u>할 수 있음

$$\hat{t}_{y,reg} = \sum_{h=1}^{H} w_k \left(\sum_{s_h}\sum_{p=1}^{P} x_k \tilde{y}_{s_{hp}} + \sum_{s_h} e_k \right)$$

· w_k : h층 i조사구 j가구 k가구원의 설계 가중값 ($= w_{hijk}$)

· s_h : h층의 표본, $\sum_{s_h} = \sum_{i=1}^{n_h}\sum_{j=1}^{20}\sum_{k=1}^{u_j}$

· n_h : h층의 표본 조사구 수

· u_j : j가구의 15세 이상 가구원 수

· p : 사후층, 성 및 연령 그룹 ($P=26$)

· $x_k = \begin{cases} 1, & k \in s_p \\ 0, & k \notin s_p \end{cases}$: 보조 정보(p사후층 여부)

· $e_k = y_k - \hat{y}_k = y_k - \tilde{y}_{s_{hp}}$: 표본 잔차

· $\hat{y}_k = x_k \hat{B}_{hp}$: 예측값

· $\hat{B}_{hp} = \dfrac{\sum_{s_{hp}} w_k \times y_k}{\sum_{s_{hp}} w_k} = \tilde{y}_{s_{hp}}$: h층, p사후층의 회귀계수 추정량

※ 회귀 추정량은 근사적으로 불편성을 만족하는 추정량으로 보조 정보가 관심 변수(y)와 선형관계에 있을 때 불편추정량보다 분산이 작아지므로, <u>분산은 회귀 추정량에 대한 선형근사 분산 추정량으로 추정</u>

○ 총계의 분산

- 테일러 전개(taylor expansion)에 근거한 선형근사(linear approximation) 분산 추정방법 사용

· 총계의 회귀 추정량의 분산은 잔차의 분산과 동일하므로 잔차의 분산을 선형근사방법으로 추정

· SAS 9.4의 Proc SurveyReg procedure를 이용하여 추정값 계산

$$\hat{V}(\hat{t}_{y,reg}) = \sum_{h=1}^{H} \hat{V}_h(\hat{t}_{y,reg}) \ , \quad \hat{V}_h(\hat{t}_{y,reg}) = \hat{V}_h(t_e)$$

$$\hat{V}_h(t_e) = w_k^2 \left[\left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (e_{i..} - \bar{e}_h)^2 + \sum_{i=1}^{n_h} \left(\frac{n_h M_{hi}}{M_h}\right) \left(1 - \frac{20}{M_{hi}}\right) \frac{20}{19} \sum_{j=1}^{20} (e_{ij.} - \bar{e}_i)^2 \right]$$

· $e_{i..} = \sum_{j=1}^{20} \sum_{k=1}^{u_j} e_{ijk}$: i조사구의 잔차합

· $\bar{e}_h = \sum_{i=1}^{n_h} e_{i..}/n_h$: h층의 조사구별 잔차합의 평균

· $e_{ij.} = \sum_{k=1}^{u_j} e_{ijk}$: i조사구 내 j가구의 잔차합

· $\bar{e}_i = \sum_{j=1}^{20} e_{ij.}/20$: i조사구의 가구당 잔차합의 평균

· N_h : h층의 모집단 조사구 수

○ 표준오차(SE) 및 상대표준오차(RSE)

$$\widehat{SE}(\hat{t}_{y,reg}) = \sqrt{\hat{V}(\hat{t}_{y,reg})}$$

$$\widehat{RSE}(\hat{t}_{y,reg}) = \frac{\widehat{SE}(\hat{t}_{y,reg})}{\hat{t}_{y,reg}} \times 100\,(\%)$$

나. 연간 추정

○ 총계 추정

$$\hat{t}_{year} = \frac{1}{12}\sum_{m=1}^{12}\hat{t}_m$$

· $\hat{t}_m = \hat{t}_{y,reg}^{m}$: m월의 총계 추정량

○ 총계의 분산 추정

$$\hat{V}(\hat{t}_{year}) = \frac{1}{144}\left[\sum_{m=1}^{12}\hat{V}\left(\hat{t}_m\right) + \sum_{m=1}^{12}\sum_{s=1,\, s\neq m}^{12}\widehat{Cov}\left(\hat{t}_m, \hat{t}_s\right)\right]$$

$$\approx c_\rho \frac{1}{12}\sum_{m=1}^{12}\hat{V}\left(\hat{t}_m\right) = c_\rho \overline{\hat{V}}\left(\hat{t}\right)$$

· $\widehat{Cov}\left(\hat{t}_m, \hat{t}_s\right) = \rho_{ms}\sqrt{Var\left(\hat{t}_m\right)\times Var\left(\hat{t}_s\right)}$

· $c_\rho = \frac{1}{144}\left(12 + \sum_{m=1}^{12}\sum_{s=1,\, s\neq m}^{12}\hat{\rho}^{|s-m|}\right)$

· $\hat{t}_m$: m월의 총계 추정량

· $\hat{V}(\hat{t}_m)$: m월 총계의 분산 추정량

· $\widehat{Cov}\left(\hat{t}_m, \hat{t}_s\right)$: m, s월 총계 추정량의 공분산

· ρ_{ms} : 자기상관계수($\hat{t}_m$과 $\hat{t}_s$의 상관계수)

* 월 추정량의 분산이 $V(\hat{t}_0)$로 동일하고 자기상관계수가 ρ인 AR(1) 모형을 따를 때, 월별 추정량의 공분산은 다음과 같이 근사함, $Cov\left(\hat{t}_m, \hat{t}_s\right) \approx \rho^{|s-m|}\, V(\hat{t}_0)$

○ 표준오차(SE) 및 상대표준오차(RSE)

$$\widehat{SE}(\hat{t}_{year}) = \sqrt{\hat{V}(\hat{t}_{year})}$$

$$\widehat{RSE}(\hat{t}_{year}) = \frac{\widehat{SE}(\hat{t}_{year})}{\hat{t}_{year}}\times 100\,(\%)$$