
2022년 이민자 체류실태 및 고용조사 표본설계 보고서

2022. 4.



〈목 차〉

I. 이민자 체류실태 및 고용조사 개요	1
II. 외국인 표본설계	3
1. 모집단 및 표본추출틀	3
2. 층화 및 분류지표	6
3. 표본 규모 및 배분	7
4. 표본 추출	9
III. 귀화허가자 표본설계	10
1. 모집단 및 표본추출틀	10
2. 층화 및 분류지표	13
3. 표본 규모 및 배분	14
4. 표본 추출	16
IV. 가중값 작성 및 추정	17
1. 가중값	17
2. 추정	19

1. 조사 목적

- 저출산·고령화 사회 진입에 따른 부족한 노동력의 국외유입으로 이들의 체류관리 및 사회통합 현황 파악
- 체류 외국인 및 귀화허가자 실태 파악을 통해 체류관리 및 사회통합 등 이민자 관련 정책수립을 위한 기초 통계자료 구축

2. 조사 연혁

- 2017년 이민자 체류실태 및 고용조사 실시(외국인 10,000명, 귀화허가자 4,000명)
 - 2012년부터 2016년까지 '외국인 고용조사'를 실시하였으나, 2017년 확대 개편(외국인과 귀화허가자 분리 등)하여 통계작성 변경 승인
 - 법적근거: 통계법 제17조, 통계법 시행령 제22조에 의한 지정통계(승인번호 920018호)
- 2018년 이민자 체류실태 및 고용조사 실시(외국인 13,500명, 귀화허가자 5,000명)
- 2019년 이민자 체류실태 및 고용조사 실시(외국인 15,000명, 귀화허가자 5,000명)
- 2020년 이민자 체류실태 및 고용조사 실시(외국인 20,000명, 귀화허가자 5,000명)

3. 조사 대상

- 조사 기간
 - 조사기준일: 2022. 5. 15.(일)
 - 조사대상기간(고용부문): 2022. 5. 15.(일) ~ 5. 21.(토)
 - 조사실시시간: 2022. 5. 24.(화) ~ 6. 7.(화)
- 조사 주기
 - 조사 주기 및 작성 주기: 매년
 - 공통항목은 매년 또는 2년 주기, 특성항목은 3년 주기
- 조사 대상: 만 15세 이상 이민자(외국인과 귀화허가자)
- 조사 단위: 이민자 개인(외국인, 귀화허가자)

4. 모집단 및 표본설계

○ 모집단

- 조사기준일(2022년 5월 15일) 현재, 대한민국에 상주*하는 만 15세 이상 외국인과 귀화허가자
 - 외국인: 91일 이상 계속 체류 또는 체류하고자 하는 외국인
 - 귀화허가자: 최근 5년 동안 대한민국으로 귀화한 귀화자(귀화허가자)
- * 90일 이하 단기 체류 외국인, 만 15세 미만의 외국인과 귀화허가자 제외

○ 표본설계

- 모집단 특성을 반영하기 위해 **외국인과 귀화허가자를 구분하여 설계**
- 표본추출틀
 - 외국인: 2022년 1월 31일 기준 법무부 등록외국인 및 외국국적동포 명부
 - * 등록면제자, 체류만료 예정자, 불법체류자, 제주 투자이민 제외
 - 귀화허가자: 법무부 귀화허가자 명부(2017. 1. 1. ~ 2022. 1. 31.)
- 표본추출방법: 층화 계통추출(Stratified Systematic Sampling)
 - '2013년 외국인 고용조사'부터 조사비용 및 업무량을 고려하여 이상추출을 도입했으나 표본 규모('13년 10,000명 → '20년 25,000명)를 고려하여 목록 추출(1상)로 변경
 - 이상추출에서 발생한 과대 가중값 문제와 이민자 특성별 포함률 차이 해소
- 표본추출단위: 이민자 개인(외국인, 귀화허가자)
- 표본규모: 전체 25,000명(외국인 20,000명, 귀화허가자 5,000명)
 - * 2020년부터 동일 규모 유지

5. 조사 항목

- 공통항목: 이민자(외국인, 귀화허가자) 전체 대상
 - 기본사항, 고용, 교육, 주거 및 생활환경, 소득과 소비, 자녀교육 등 8개 부문
- 특성항목: 귀화허가자, 영주(F-5), 국민의 배우자(F-5-2) 체류자격 소지자 대상
 - 체류자격에 따라 관련 정책에 필요한 세부 항목

6. 조사 방법

- 조사원에 의한 방문면접조사와 자기기입식 인터넷 조사(유학생, 전문인력) 병행

1. 모집단 및 표본추출틀

가. 목표모집단

- 조사기준일(2022. 5. 15.) 현재, 대한민국에 상주*하는 만 15세 이상 외국인
* 90일 이하 단기 체류 외국인 및 만 15세 미만의 외국인 제외

나. 조사모집단

- 2022년 1월 31일 기준 법무부 외국인 등록명부 및 국내거소신고 외국인 동포 명부에 포함된 만 15세 이상 외국인
- 등록면제자, 체류만료 예정자, 불법체류자, 제주 투자이민 제외

<표 1> 2022년 외국인 모집단 현황(2022. 1. 31. 기준)

(단위: 명)

전체	조사 모집단	제외자 소계								
			외교 (A1)	공무 (A2)	협정 (A3)	부동산투자 (F2-8)*	부동가족 (F-2-81)*	"00" 체류자격	체류기간 만료자	귀화자 중복
1,490,860	1,331,122	159,738	3,680	2,326	7,980	342	433	13,788	130,900	289

* 주소지가 제주도인 경우만 제외

다. 표본추출틀

- 현장조사 효율성을 고려하여 조사모집단에서 과소 행정동 및 사무소(분소) 내 외국인을 제외한 등록외국인 및 외국국적동포 명부로 구축
- 과소 행정동 기준: 행정동 내 외국인 수 100명 미만
- 과소 사무소(분소) 기준: 관할 구역 내 외국인 수 0.15%(약 2,000명) 미만
* 사무소(분소)별 귀화허가자 분포를 고려하여 결정

라. 포함률

- 조사모집단 대비 표본추출틀 **포함률 95.9%**
- 전체 1,331,122명 중 1,275,950명 포함(제외 55,172명)

- 8개 권역(공포 수준)별 포함률은 78.1~99.8% 수준으로 나타남
 - 권역 및 시도별 포함률 차이를 고려하여 지역을 층화변수로 설정
- 성, 연령, 체류자격 및 국적별 외국인 포함률은 평균 94.6%로 높게 나타남
- 외국인을 추출하는 목록 추출틀의 포함률 제고*를 위해 표본추출방법을 이상(two phase)에서 일상(one phase)으로 변경
 - * '21년 2상(2nd phase) 표본추출틀의 포함률: 전국 78.3%, 시도별 41.6%~87.6%, 국적별 66.7%~83.8%
- 외국인 특성별 분포도 모집단과 유사하여 표본추출틀이 모집단을 잘 대표하는 것으로 판단

<표 2> 거주 지역별 모집단 및 표본추출틀 분포

(단위: 명, %)

거주 지역 (8개 권역 및 시도)	조사모집단		표본추출틀*		포함률 (B/A*100)
	외국인(A)	구성비	외국인(B)	구성비	
전국	1,331,122	100.0	1,275,950	100.0	95.9
서울	323,150	24.3	322,387	25.3	99.8
인천	83,513	6.3	82,757	6.5	99.1
경기	490,644	36.9	488,423	38.3	99.5
부산·울산·경남	115,050	8.6	105,746	8.3	91.9
부산	36,647	2.8	34,855	2.7	95.1
울산	19,444	1.5	19,243	1.5	99.0
경남	58,959	4.4	51,648	4.1	87.6
대전·충남·충북·세종	141,075	10.6	131,436	10.3	93.2
대전	17,955	1.4	16,717	1.3	93.1
충남	75,267	5.7	71,682	5.6	95.2
충북	42,813	3.2	38,429	3.0	89.8
세종	5,040	0.4	4,608	0.4	91.4
대구·경북	68,510	5.2	58,157	4.6	84.9
대구	24,342	1.8	22,559	1.8	92.7
경북	44,168	3.3	35,598	2.8	80.6
광주·전남·전북	74,925	5.6	58,522	4.6	78.1
광주	20,941	1.6	19,727	1.6	94.2
전남	27,184	2.0	17,789	1.4	65.4
전북	26,800	2.0	21,006	1.7	78.4
강원·제주	34,255	2.6	28,522	2.2	83.3
강원	16,444	1.2	11,146	0.9	67.8
제주	17,811	1.3	17,376	1.4	97.6

* 과소 행정동 및 사무소(분소) 내 외국인 제외

<표 3> 외국인 특성별 모집단 및 표본추출률 분포

(단위: 명, %)

외국인 특성		조사모집단		표본추출률*		포함률 (b)/a*100
		외국인(a)	구성비	외국인(b)	구성비	
전국		1,331,122	100.0	1,275,950	100.0	95.9
성	남	696,479	52.3	667,990	52.4	95.9
	여	634,643	47.7	607,960	47.7	95.8
연령	29세 이하	296,683	22.3	281,832	22.1	95.0
	30-39세	360,210	27.1	343,083	26.9	95.2
	40-49세	213,132	16.0	205,151	16.1	96.3
	50-59세	240,626	18.1	232,731	18.2	96.7
	60세 이상	220,471	16.6	213,153	16.7	96.7
체류자격	비전문취업	162,683	12.2	146,206	11.5	89.9
	방문취업	116,907	8.8	115,368	9.0	98.7
	전문인력	38,498	2.9	37,083	2.9	96.3
	유학생	123,917	9.3	122,751	9.6	99.1
	재외동포	470,871	35.4	460,896	36.1	97.9
	영주	156,108	11.7	151,605	11.9	97.1
	결혼이민	128,488	9.7	116,284	9.1	90.5
	동반	60,942	4.6	56,442	4.4	92.6
	기타	72,708	5.5	69,315	5.4	95.3
국적	중국(비한국계)	135,057	10.2	131,148	10.3	97.1
	중국(한국계)	578,921	43.5	570,092	44.7	98.5
	베트남	121,216	9.1	109,260	8.6	90.1
	필리핀	28,655	2.2	26,255	2.1	91.6
	인도네시아	23,318	1.8	20,996	1.7	90.0
	타이	26,205	2.0	24,017	1.9	91.7
	우즈베키스탄	49,736	3.7	48,563	3.8	97.6
	몽골	16,441	1.2	16,031	1.3	97.5
	일본	23,837	1.8	21,998	1.7	92.3
	캄보디아	31,527	2.4	27,519	2.2	87.3
	네팔	30,767	2.3	26,458	2.1	86.0
	기타아시아	110,231	8.3	104,614	8.2	94.9
	미국	59,714	4.5	56,432	4.4	94.5
	캐나다	19,774	1.5	18,797	1.5	95.1
	오세아니아	7,974	0.6	7,554	0.6	94.7
유럽	53,130	4.0	51,989	4.1	97.9	
기타	14,619	1.1	14,227	1.1	97.3	

* 과소 행정동 및 사무소(분소) 내 외국인 제외

2. 총화 및 분류지표

가. 총화

- 조사모집단 분석 결과, 시도별로 외국인의 체류자격과 국적 분포에 차이가 있으며, 표본추출틀의 포함률 차이도 있으므로 시도별 특성을 반영하기 위해 17개 시도로 총화
- 외국인의 체류자격은 '재외동포'가 35.4%로 가장 많고, '비전문취업' 12.2%, '영주' 11.7% 순으로 높게 나타남
 - '재외동포'는 서울(42.2%)과 경기(41.1%)에서 상대적으로 높게 나타났고, '비전문취업'은 경남(29.5%)과 전남(29.2%), '영주'는 제주(16.7%)와 인천(15.9%)에서 다른 지역보다 높은 비중을 차지하는 것으로 나타남
- 외국인의 국적은 '한국계 중국'이 43.5%로 가장 많고, '기타 아시아' 25.6%, '중국' 10.2%로 높게 나타남
 - '한국계 중국'은 서울(55.3%)과 경기(54.5%)에서 상대적으로 높게 나타났고, '비한국계 중국'은 제주(27.4%)와 대전(19.2%)에서 높게 나타남
 - '베트남' 국적은 전라남북도와 대구·경북 지역에서 20% 이상 높은 비중을 차지하고 있으며, '기타 아시아'는 경상남북도와 전남, 충청에서 40% 이상으로 높게 나타남
 - 비중이 낮은 북미 출신 외국인인 서울, 대전, 세종에서 다른 지역보다 높게 나타났으며, 유럽은 부산과 광주, 기타 국적은 대전과 세종에서 높게 나타남
- 외국인의 체류자격, 국적 등의 특성을 고려한 설계 필요

※ 17개 시도별 외국인 체류자격 및 국적 분포는 미공표 통계이므로 첨부하지 않음

나. 분류지표

- 시도별 체류자격과 국적의 분포에 차이가 있으므로, 설계층인 17개 시도 내에서 체류자격과 국적을 분류지표로 활용
 - 체류자격 대대분류와 세분류를 1차와 3차 분류지표로 각각 사용하여 표본 내 세분류 수준에서의 체류자격이 편중되지 않도록 고려
 - 특히, 주요 정책관리 대상인 비전문취업(E-9) 내 세분류 분포 반영
 - 국적은 기타 아시아와 북미 국가의 국적으로 세분한 17개 범주로 구분

<표 4> 분류지표

순서	분류지표	
1차	□ 체류자격(대대분류)	
	① 비전문취업 ② 방문취업 ③ 전문인력	
	④ 유학생 ⑤ 재외동포 ⑥ 영주자	
	⑦ 결혼이민자 ⑧ 동반 ⑨ 기타	
2차	□ 국적	
	① 기타 중국 ② 한국계 중국 ③ 베트남	
	④ 필리핀 ⑤ 인도네시아 ⑥ 타이	
	⑦ 우즈베키스탄 ⑧ 몽골 ⑨ 일본	
	⑩ 캄보디아 ⑪ 네팔 ⑫ 기타 아시아	
	⑬ 미국 ⑭ 캐나다 ⑮ 오세아니아	
	⑯ 유럽 ⑰ 기타 국가	
	3차	□ 체류자격(세분류)
		- 153개 세분류

3. 표본 규모 및 배분

가. 표본 규모

- 외국인의 전국 표본규모는 '21년과 동일하게 20,000명 유지

<표 5> 외국인 표본규모 및 상대표준오차

구분	2017년	2018년	2019년	2020년	2021년
표본 규모	10,000	13,500	15,000	20,000	20,000
취업자 RSE	0.9	0.8	0.8	0.7	0.7
실업자 RSE	7.2	6.2	5.4	3.9	4.4

(단위: 명, %)

나. 표본 배분

- 시도별 외국인 수의 제공근에 비례하도록 배분한 후, 전년도 표본규모와 취업자 및 실업자의 예상 상대표준오차(RSE)를 고려하여 조정
 - 실업자의 상대표준오차가 상대적으로 큰 인천과 충청권, 강원·제주 지역의 표본규모 확대
 - 인천 1,465명 → 1,600명, 충남 1,341명 → 1,400명, 세종 121명 → 300명, 강원 663명 → 800명, 제주 709명 → 800명

<표 6> 외국인 표본 배분 결과

(단위: 명, %)

설계층 ¹⁾	2021년 조사결과				제공근 비례배분			조정안			
	표본규모		RSE		제공근 구성비	표본 규모		표본 규모		예상 RSE ²⁾	
	시도	권역	취업자	실업자		시도	권역(계)	시도	권역(계)	취업자	실업자
전국	20,000	20,000	0.7	4.4	1.000	20,000	20,000	20,000	20,000	0.7	4.4
서울	2,962	2,962	1.9	8.7	0.145	2,901	2,901	2,800	2,800	1.9	8.9
인천	1,465	1,465	2.5	14.4	0.074	1,475	1,475	1,600	1,600	2.4	13.8
경기	3,591	3,591	1.3	8.9	0.179	3,574	3,574	3,400	3,400	1.3	9.1
부산	996				0.049	977		950			
울산	735	3,019	1.6	10.4	0.036	712	2,928	700	2,850	1.7	10.7
경남	1,288				0.062	1,239		1,200			
대전	683				0.034	684		700			
충남	1,341	3,208	1.5	12.7	0.070	1,400	3,502	1,400	3,400	1.4	12.4
충북	1,063				0.053	1,056		1,000			
세종	121				0.018	362		300			
대구	820	1,918	1.9	11.9	0.040	796	1,869	800	1,900	1.9	11.9
경북	1,098				0.054	1,072		1,100			
광주	742				0.037	738		750			
전남	862	2,465	1.8	10.8	0.042	841	2,415	850	2,450	1.8	10.8
전북	861				0.042	835		850			
강원	663	1,372	2.6	14.8	0.033	654	1,336	800	1,600	2.4	13.7
제주	709				0.034	681		800			

1) 공표 단위인 8개 권역 내에서 시도 구분

2) 예상 상대표준오차: 전년도 표본규모와 상대표준오차를 이용하여 산출

$$RSE_{new} = \sqrt{\frac{n_{old}}{n_{new}}} \times RSE_{old}$$

RSE_{new} : 예상 상대표준오차 n_{new} : 새로운 표본규모
 RSE_{old} : '21년 상대표준오차 n_{old} : '21년 표본규모

4. 표본 추출

- 표본 추출 방법: 층화 계통추출(Stratified Systematic Sampling)
 - 설계층: 17개 시도
 - 분류지표: 체류자격 및 국적
- 표본 추출 시 고려사항
 - 조사관리 및 업무효율을 고려하여 사무소별 최소 30명* 이상 추출
 - * '21년 기준, 조사원 1인당 평균 33명 조사(외국인+귀화허가자)
 - 비전문취업(E-9)의 경우 체류자격 세분류 분포 고려
- 예비 표본 추출
 - 원 표본과 1:1로 매칭되도록 2배수의 예비(대체) 표본 추출
 - 원 표본과 대체 표본의 주소를 고려하여 추출
 - 동일한 대체 사유가 발생하지 않도록 가능한 한 다른 주소지로 대체
 - 조사원의 이동 거리가 증가하지 않도록 인접한 지역에서 추출(건물관리 번호 및 행정구역코드 고려)
 - 표본의 특성을 고려하여 추출
 - 특성항목 조사대상(영주권자^{F-5}, 국민의 배우자^{F-5-2})은 동일한 체류자격과 국적을 소지한 자로 대체

1. 모집단 및 표본추출틀

가. 목표모집단

- 최근 5년 이내* 한국 정부로부터 귀화 허가를 받은 만 15세 이상 귀화허가자로 조사기준일(2022. 5. 15.) 현재 비상주인구**가 아닌 자

* 2017년 1월 1일 이후 귀화허가자

** 비상주인구: 조사기준일(5월 15일) 현재 국내 거주하지 않으며, 비거주기간이 연속 91일 이상인 자

나. 조사모집단

- 2017년 1월 1일부터 2022년 1월 31일 사이에 한국 정부에서 귀화 허가를 받은 만 15세 이상 귀화허가자

- 재외국민¹⁾, 거주불명자²⁾, 국적상실자 제외

1) 재외국민: 2021년 11월 1일 현재 한국에 거주하지 않는 귀화자

2) 거주불명자: 2021년 11월 1일 현재 등록기준지에 거주하지 않는 귀화자

* 인구총조사과의 자료와 연계되는 귀화자는 주민등록상의 등록기준지를 사용하고 그 외에 연계되지 않은 9,552명은 법무부의 체류지 주소로 대체

<표 7> 2022년 귀화허가자 모집단 현황(2022. 1. 31. 기준)

(단위: 명)

전체	조사모집단	제외자 소계	주소 미상	재외국민	거주불명자
50,704	50,671	33	3	1	29

다. 표본추출틀

- 현장조사 효율성을 고려하여 조사모집단에서 과소 행정동 및 사무소(분소) 내 귀화허가자를 제외한 귀화허가자 명부로 구축

- 과소 행정동 기준: 행정동 내 귀화허가자 수 3명 미만

- 과소 사무소(분소) 기준: 관할 구역 내 귀화허가자 수 150명 미만

* 사무소(분소)별 외국인 분포를 고려하여 결정

라. 포함률

○ 조사모집단 대비 표본추출률 **포함률 95.7%**

- 전체 50,671명 중 48,473명 포함(제외 2,198명)
- 8개 권역(공표 수준)별 포함률은 84.6~99.4% 수준으로 나타남
 - 권역 및 시도별 포함률 차이를 고려하여 지역을 층화변수와 분류지표로 고려
- 성, 연령, 혼인상태 및 이전 국적별 포함률은 평균 94.5%로 높게 나타남

<표 8> 거주 지역별 귀화허가자 모집단 및 표본추출률 분포

(단위: 명, %)

거주 지역 (8개 권역 및 시도)	조사모집단		표본추출률*		포함률 (㉔/㉓*100)
	귀화허가자(㉓)	구성비	귀화허가자(㉔)	구성비	
전국	50,671	100.0	48,473	100.0	95.7
서울	9,339	18.4	9,097	18.8	97.4
인천	4,002	7.9	3,913	8.1	97.8
경기	15,703	31.0	15,505	32.0	98.7
부산·울산·경남	5,389	10.6	5,026	10.4	93.3
부산	1,851	3.7	1,739	3.6	93.9
울산	935	1.9	921	1.9	98.5
경남	2,603	5.1	2,366	4.9	90.9
대전·충남·충북·세종	5,445	10.8	5,149	10.6	94.6
대전	937	1.9	918	1.9	98.0
충남	2,634	5.2	2,577	5.3	97.8
충북	1,651	3.3	1,434	3.0	86.9
세종	223	0.4	220	0.5	98.7
대구·경북	3,806	7.5	3,506	7.2	92.1
대구	1,390	2.7	1,318	2.7	94.8
경북	2,416	4.8	2,188	4.5	90.6
광주·전남·전북	4,838	9.6	4,332	8.9	89.5
광주	1,008	2.0	963	2.0	95.5
전남	2,071	4.1	1,770	3.7	85.5
전북	1,759	3.5	1,599	3.3	90.9
강원·제주	2,149	4.2	1,945	4.0	90.5
강원	1,291	2.6	1,092	2.3	84.6
제주	858	1.7	853	1.8	99.4

* 과소 행정동 및 사무소(분소) 내 귀화허가자 제외

- 귀화허가자를 추출하는 목록 추출들의 포함률 제고*를 위해 표본추출방법을 이상(two phase)에서 일상(one phase)으로 변경

* '21년 2상(2nd phase) 표본추출들의 포함률: 전국 73.0%, 시도별 34.8%~94.7%, 국적별 63.4%~83.3%

- 성, 연령, 혼인상태 및 이전 국적별 표본추출들의 분포를 살펴본 결과, 모집단과 표본추출들의 분포가 유사하여 모집단을 잘 대표하는 것으로 판단

<표 9> 귀화허가자 특성별 모집단 및 표본추출률 분포

(단위: 명, %)

외국인 특성		조사모집단		표본추출률*		포함률 (b/a*100)
		귀화허가자(a)	구성비	귀화허가자(b)	구성비	
전국		50,671	100.0	48,473	100.0	<u>95.7</u>
성	남	9,275	18.3	8,932	18.4	96.3
	여	41,396	81.7	39,541	81.6	95.5
연령	29세 이하	10,880	21.5	10,130	20.9	93.1
	30-39세	21,472	42.4	20,682	42.7	96.3
	40-49세	8,268	16.3	8,026	16.6	97.1
	50-59세	5,099	10.1	4,935	10.2	96.8
	60세 이상	4,952	9.8	4,700	9.7	94.9
혼인상태	미혼/미상	29,269	57.8	27,956	57.7	95.5
	기혼	21,402	42.2	20,517	42.3	95.9
이전 국적	중국(비한국계)	5,319	10.5	5,066	10.5	95.2
	중국(한국계)	16,979	33.5	16,551	34.1	97.5
	베트남	20,400	40.3	19,311	39.8	94.7
	캄보디아	1,721	3.4	1,613	3.3	93.7
	필리핀	1,925	3.8	1,820	3.8	94.5
	타이완	1,494	3.0	1,419	2.9	95.0
	몽골	483	1.0	468	1.0	96.9
	기타 아시아	1,759	3.5	1,661	3.4	94.4
비아시아	591	1.2	564	1.2	95.4	

* 과소 행정등 및 사무소(분소) 내 귀화허가자 제외

2. 총화 및 분류지표

가. 총화

- 조사모집단 분석 결과, 지역별로 귀화허가자의 성, 연령, 혼인상태 및 이전 국적 분포의 차이가 나타남
- 시도별 모집단 크기가 작으므로 공표 수준인 8개 권역*으로 지역층 설정
 - * 권역: 서울, 인천, 경기, 부산·울산·경남, 대전·충남·충북·세종, 대구·경북, 광주·전남·전북, 강원·제주
- 권역 내에서 시도 구분을 1차 분류지표로 활용하여 시도 분포 반영
- 남성 귀화허가자의 안정적인 통계작성을 위해 성별을 총화변수로 적용
- 전국을 8개 권역과 성별에 따라 16개 그룹으로 총화
- 귀화허가자의 국적은 '베트남'이 40.3%로 가장 많고, 다음으로 '한국계 중국'이 33.5%로 높게 나타남
 - '남성'의 이전 국적은 '한국계 중국'이 63.4%로 가장 많고, '여성'의 이전 국적은 '베트남'이 48.3%로 가장 많음
 - 지역별로는 수도권에서 '한국계 중국'의 비중이 상대적으로 높게 나타났으며, 전라도와 경상도 지역은 '베트남' 비중이 높게 나타남
- 귀화허가자의 성, 연령계층, 이전 국적 등의 특성을 고려한 설계 필요

※ 17개 시도별 귀화허가자 특성 분포는 미공표 통계이므로 첨부하지 않음

나. 분류지표

- 설계 층(8개 권역 및 성별) 내에서 이전 국적, 혼인상태, 연령계층을 공통 분류지표로 사용
- 비수도권의 5개 권역은 시도 구분을 1차 분류지표로 적용하고, 수도권의 3개 권역은 마지막 분류지표로 시군구 적용

<표 10> 귀화허가자 분류지표

순서	분류지표
(1차)	□ 시도 (비수도권)
2차	□ 이전 국적 ① 기타 중국 ② 한국계 중국 ③ 베트남 ④ 캄보디아 ⑤ 필리핀 ⑥ 타이완 ⑦ 몽골 ⑧ 기타 아시아 ⑨ 기타 국가
3차	□ 혼인상태 ① 미혼/미상 ② 기혼(유배우자, 사별, 이혼)
4차	□ 연령계층 ① 15~29세 ② 30~39세 ③ 40~49세 ④ 50~59세 ⑤ 60세 이상
(5차)	□ 시군구 (서울, 인천 경기)

3. 표본 규모 및 배분

가. 표본 규모

- 귀화허가자의 전국 표본규모는 '21년과 동일하게 5,000명 유지

<표 11> 귀화허가자 표본규모 및 상대표준오차

(단위: 명, %)

구분	2017년	2018년	2019년	2020년	2021년
표본 규모	4,000	5,000	5,000	5,000	5,000
취업자 RSE	1.5	1.3	1.3	1.4	1.2
실업자 RSE	9.9	8.9	8.2	6.8	7.9

나. 표본 배분

- 성별 표본은 전년과 동일하게 남성 1,500명, 여성 3,500명으로 유지
- 성별 층 내에서 8개 권역별 귀화허가자 수의 제공근에 비례하도록 배분한 후, 전년도 표본규모와 예상 상대표준오차(RSE)를 고려하여 조정
- 상대표준오차가 상대적으로 안정적인 경기의 표본을 축소하고, 서울(남)과 충청권(여)의 표본규모 확대

<표 12> 귀화허가자 표본 배분 결과

(단위: 명, %)

설계층	2021년 조사결과				제공근 비례배분			조정안		
	표본 규모	RSE		귀화허가자 (조사모집단)	제공근 구성비	표본 규모	표본 규모	예상 RSE		
		취업자	실업자*					취업자	실업자*	
남	소계	1,500	3.0	15.9	9,275	1.000	1,500	1,500	3.0	15.9
	서울	332	6.1		2,629	0.214	321	340	6.0	
	인천	199	7.1		1,025	0.134	200	200	7.1	
	경기	378	5.1		3,695	0.254	380	360	5.2	
	부산·울산·경남	143	9.8		522	0.095	143	150	9.5	
	대전·충남·충북·세종	161	8.4		669	0.108	162	150	8.7	
	대구·경북	103	12.2		273	0.069	103	110	11.8	
	광주·전남·전북	98	11.2		263	0.068	102	100	11.1	
	강원·제주	86	12.5		199	0.059	88	90	12.3	
여	소계	3,500	1.8	9.1	41,396	1.000	3,500	3,500	1.8	9.1
	서울	518	5.1		6,710	0.147	515	520	5.1	
	인천	348	5.6		2,977	0.098	343	350	5.6	
	경기	685	3.9		12,008	0.197	690	640	4.0	
	부산·울산·경남	446	4.6		4,867	0.125	439	440	4.6	
	대전·충남·충북·세종	431	4.9		4,776	0.124	435	470	4.6	
	대구·경북	379	4.7		3,533	0.107	374	380	4.7	
	광주·전남·전북	417	4.6		4,575	0.122	426	420	4.6	
	강원·제주	276	5.2		1,950	0.079	278	280	5.1	
전국		5,000	1.2	7.9	50,671		5,000	5,000	1.2	7.9
권역	서울	850	3.1	14.4	9,339		836	860	3.1	14.3
	인천	547	3.4	23.5	4,002		544	550	3.4	23.5
	경기	1,063	2.4	16.7	15,703		1,070	1,000	2.5	17.2
	부산·울산·경남	589	3.7	22.9	5,389		582	590	3.7	22.8
	대전·충남·충북·세종	592	3.7	31.6	5,445		597	620	3.6	30.9
	대구·경북	482	4.0	24.6	3,806		477	490	4.0	24.4
	광주·전남·전북	515	4.0	28.2	4,838		527	520	4.0	28.0
	강원·제주	362	4.2	29.8	2,149		366	370	4.1	29.4

* 실업자는 권역 내에서 성별로 공표하지 않음

4. 표본 추출

- 표본 추출 방법: 층화 계통추출(Stratified Systematic Sampling)
 - 설계층: 성과 8개 권역*으로 구분하여 16개 층 설정
 - * 권역: 서울, 인천, 경기, 부산·울산·경남, 대전·충남·충북·세종, 대구·경북, 광주·전남·전북, 강원·제주
 - 분류지표: 행정구역(시도/시군구), 이전 국적, 혼인상태, 연령계층
- 예비 표본 추출
 - 원 표본과 1:1로 매칭되도록 1배수의 예비(대체) 표본 추출
 - 원 표본과 대체 표본의 주소를 고려하여 추출
 - 동일한 대체 사유가 발생하지 않도록 가능한 한 다른 주소지로 대체
 - 조사원의 이동 거리가 증가하지 않도록 인접한 지역에서 추출(건물관리 번호 및 행정구역코드 고려)

1. 가중값

- (설계 가중값) 설계층별 추출률의 역수를 이용하여 작성
 - 외국인과 귀화허가자 표본을 구분하여 설계하였으므로, 각각의 표본추출 방법에 맞게 작성
 - 외국인: 특·광역시 및 시도로 구분한 17개 지역층 적용
 - 귀화허가자: 성 및 8대 권역으로 구분한 16개 층 적용

$$\text{설계 가중값} \quad w_{hi}^b = \frac{N_h}{n_h}$$

- $h = 1, 2, \dots, H$: 설계층
- $i = 1, 2, \dots, n_h$: h 층의 표본
- w_{hi}^b : h 층 i 표본의 설계 가중값
- N_h : h 층의 모집단 이민자 수(표본추출률 기준)
- n_h : h 층의 표본 이민자 수

- (무응답 가중값) 응답성향점수를 이용하여 무응답 조정계수 산출
 - 이민자의 특성(성, 연령, 체류자격 및 국적 등)을 독립변수로 하는 로지스틱 회귀모형을 이용하여 응답성향점수(응답확률) 추정
 - 외국인: 성별, 연령계층, 국적, 체류자격 활용
 - 귀화허가자: 연령계층, 이전 국적, 거주 지역 활용

$$\text{무응답 가중값} \quad w_{hi}^r = w_{hi}^b \times \frac{1}{\hat{p}_{hi}}$$

- $h = 1, 2, \dots, H$: 설계층
- $i = 1, 2, \dots, n_h^r$: h 층의 응답 표본
- w_{hi}^b : h 층 i 표본의 설계 가중값
- \hat{p}_{hi} : h 층 i 표본의 응답성향점수(응답확률)

- (사후 가중값) 최신 모집단의 분포에 맞게 반복비례가중법(Iterative proportional weighting/Raking ratio method)을 적용하여 최종 가중값 작성
 - 국적, 체류자격 등 공표 수준을 고려하여 외국인은 4차원, 귀화허가자는 3차원 분포에 맞춰 보정
 - 외국인: ①성(2)×국적(16) → ②성(2)×체류자격(8) → ③성(2)×권역(8) → ④연령계층(5)×체류자격(6)
 - 귀화허가자: ①이전 국적(8) → ②성(2)×권역(8) → ③성(2)×연령계층(5)
 - 반복비례 알고리즘을 이용하여 수렴 조건을 만족할 때까지 가중값 보정을 반복하고, 최종 가중값을 사후 가중값으로 사용

사후 가중값 $w_{hi} = w_{hi}^T$

- w_{hi} : h 층 i 표본의 사후 가중값
- w_{hi}^T : h 층 i 표본의 T 번째 반복비례 가중값

4차원 제약식(외국인 벤치마크 목표)

$$N_{a...} = \hat{N}_{a...} \quad N_{b..} = \hat{N}_{b..} \quad N_{..c} = \hat{N}_{..c} \quad N_{...d} = \hat{N}_{...d}$$

- a, b, c, d : 벤치마크 차원
- N_{abcd} : 모집단 외국인 수
- \hat{N}_{abcd} : 모집단 외국인 수 추정값(무응답 가중값 적용)

* 귀화허가자는 3차원 제약식 적용

수렴 조건 - 수렴 조건을 만족할 때까지 반복 수행

$$\sum_a \left| \hat{N}_{a...}^{(t)} - N_{a...} \right| + \sum_b \left| \hat{N}_{b..}^{(t)} - N_{b..} \right| + \sum_c \left| \hat{N}_{..c}^{(t)} - N_{..c} \right| + \sum_d \left| \hat{N}_{...d}^{(t)} - N_{...d} \right| < \varepsilon (= 0.000001)$$

- $\hat{N}_{abcd}^{(t)}$: 모집단 외국인 수에 대한 t 번째 추정값

* 귀화허가자는 3차원 수렴 조건 적용

2. 추정

- 점 추정: H-T(Horvitz-Thompson) 추정량

$$\text{총계 추정량} \quad \hat{\tau} = \hat{Y} = \sum_h \sum_i w_{hi} y_{hi}$$

$$\text{평균 추정량} \quad \hat{\mu} = \hat{y} = \hat{Y} / w_{..}$$

- w_{hi} : h 층 i 표본의 사후 가중값
- y_{hi} : h 층 i 표본의 관측값
- $w_{..}$: 모집단 이민자 수(= $\sum_h \sum_i w_{hi}$)

- 분산 추정: 테일러 선형근사 추정량

$$\text{총계의 분산} \quad \widehat{Var}(\hat{\tau}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (w_{hi}y_{hi} - \overline{wy}_h)^2$$

$$\text{평균의 분산} \quad \widehat{Var}(\hat{\mu}) = \widehat{Var}(\hat{\tau}) / w_{..}^2$$

$$\cdot \overline{wy}_h = \sum_{i=1}^{n_h} w_{hi}y_{hi} / n_h$$

- 표준오차(SE) 및 상대표준오차(RSE)

$$\text{표준오차} \quad \widehat{SE}(\hat{\theta}) = \sqrt{\widehat{Var}(\hat{\theta})}$$

$$\text{상대표준오차} \quad \widehat{RSE}(\hat{\theta}) = \frac{\widehat{SE}(\hat{\theta})}{\hat{\theta}} \times 100(\%)$$