

# 2019년 지역별 고용조사 자치구(군) 표본설계 보고서

(2019.2, 표본과)

## 1. 모집단 및 표본추출률

### 가. 목표모집단

- 조사기간 동안 서울특별시 및 6개 광역시\*에 거주하고 있는 만 15세 이상 모든 사람

\* 부산, 대구, 인천, 광주, 대전, 울산

- 아파트조사구(A) 거주 비중이 과반수(52.8%)를 초과하고, 시설조사구(4) 거주 비중(0.03%)이 가장 낮음

< 특·광역시 조사구특성별 분포 >

(단위 : 조사구, 명, %)

조사구특성	조사구수	전체 인구수	만15세 이상 인구	
			인구	비중
특·광역시 전체	308,882	22,443,166	19,601,681	100.00
아파트(A)	149,226	12,340,785	10,349,453	52.80
보통(1)	159,136	10,049,607	9,201,631	46.94
섬(2)	225	15,361	14,125	0.07
기숙(3)	153	30,739	30,566	0.16
시설(4)	142	6,674	5,906	0.03

### 나. 조사모집단

- 2016년 인구주택총조사(이하 인총) 결과 서울특별시 및 6개 광역시의 만 15세 이상 모든 사람
  - 군대병영, 병원, 교도소 등 특수사회시설조사구(4)를 제외한 일반조사구(1, 2, 3, A) 내 만 15세 이상 모든 사람
- 목표모집단에 대한 조사모집단의 포함률은 조사구 기준 99.95%, 만 15세 이상 인구 기준 99.97%로 시현

< 조사모집단 포함률 >

(단위 : %)

	조사구	전체 인구	≥만15세 인구
			비중
전체	99.95	99.97	99.97

## 다. 표본추출틀

- 2016년 인총 특·광역시 전수조사구 중 일반 조사구(1, 2, A)를 대상으로 등록 및 표본조사 결과를 이용하여 작성
  - 실제 조사하기 어려운 기숙 및 시설조사구(3,4) 제외
  - 자치구(군) 내 섬조사구 비중이 5% 미만인 경우 섬조사구 제외
  - 섬비중이 60%를 초과하는 인천 옹진군(61.8%)은 포함
  - 표본추출틀의 포함률은 만 15세 이상 인구 99.83% 수준임

### < 표본추출틀의 지역별 포함률 >

(단위 : 개, 명, %)

	조사모집단		표본추출틀			
	조사구	만 15세 이상	조사구	포함률	만 15세 이상	포함률
특광역시	308,740	19,595,775	308,530	99.93	19,562,175	99.83
서울	134,917	8,523,209	134,869	99.96	8,516,373	99.92
부산	48,071	2,985,945	48,052	99.96	2,980,544	99.82
대구	32,696	2,102,857	32,672	99.93	2,097,484	99.74
인천	37,094	2,459,720	37,020	99.80	2,449,356	99.58
광주	19,787	1,256,965	19,773	99.93	1,255,037	99.85
대전	20,753	1,294,132	20,735	99.91	1,292,390	99.87
울산	15,422	972,947	15,409	99.92	970,991	99.80

\* 포함률 : 모집단 대비 표본추출틀 내 인구 비중

## 2. 총화 및 분류지표

### 가. 개요

#### ◦ 총화변수

- 7개 특광역시 내 자치구 및 군을 이용한 74개 지역층을 설계층으로 사용

※ 표본추출틀 내 행정구역은 2016년 11월 1일 기준임

#### ◦ 분류지표

- 조사항목과 연관성이 높은 특성 변수(분류지표)에 따라 조사구를 정렬하고 계통추출하면, 내재적 총화효과를 얻을 수 있음
- 2015년 인총자료\*와 2018년 상반기 지역별고용조사 결과 자료를 연결하여 분류지표 선정을 위한 분석 수행

\* 2018년 지역별 고용조사 표본의 표본추출틀 정보임

## 나. 층화 분류지표 적용을 위한 군집분석

- 특·광역시는 대도시 지역으로 인구가 밀집되어 있어, 특·광역시 내 개별 자치구(설계층) 간 유사성이 있는 것으로 확인
  - 자치구(군)가 10개 이상인 서울, 부산, 인천은 자치구(군)를 군집 분석 한 후 군집별 분류지표 탐색\*

\* 대구, 광주, 대전 및 울산은 광역시별 분류지표 탐색

< 군집분석 결과(서울, 부산, 인천) >

군집	서울	부산	인천
1	용산구, 성동구, 광진구, 금천구, 영등포구, 관악구	북구, 사하구, 강서구, 사상구	중구, 동구, 남구, 서구, 연수구, 남동구, 부평구, 계양구
2	종량구, 강북구, 도봉구, 은평구, 구로구	중구, 서구, 동구, 영도구	강화군, 옹진군
3	노원구, 양천구, 강서구, 송파구, 강동구	동래구, 해운대구, 연제구, 수영구, 기장군	
4	마포구, 서초구, 강남구	부산진구, 남구, 금정구	
5	종로구, 중구, 동대문구, 성북구, 서대문구, 동작구		

## 다. 1차 분류지표

- (행정구역) 동/읍면으로 구성된 자치구(군)의 행정구역별 고용률 차이를 검정한 결과,
  - 74개 자치구(군) 모두에서 동/읍면별 고용률에 유의미한 차이가 있는 것으로 확인

☞ 모집단 분석 결과, 동읍면의 고용 특성에 유의미한 차이가 있으므로 1차 분류지표로 행정구역 구분 활용

## 라. 2차 분류지표

- (주택유형) 주택유형\*에 따른 고용률에 차이가 있는지 여부를 확인하기 위하여 분산분석 수행

\* 주택유형은 가구주 조사결과 자료를 이용하여 분석

< 주택유형 구분 >

구분	주택구분				
Type1	①단독	②소형APT*	③중형APT*	④대형APT*	⑤기타
Type2	①단독	②소형APT	③중대형APT	④기타	
Type3	①단독	②중소형APT	③대형APT	④기타	
Type4	①아파트	②그외			
Type5	①단독	②그외			
Type6	분류없음				

\* 소형APT(전용면적 60m<sup>2</sup> 이하), 중형APT(전용면적 61~102m<sup>2</sup>), 대형APT(전용면적 103m<sup>2</sup> 이상)

- 주택유형을 4개(단독, 아파트(소형·중대형), 기타)로 분류했을 때 고용률 차이가 가장 잘 나타나는 것으로 확인

< 주택유형에 따른 고용률 차이가 유의미한 설계층 수 >

(단위 : 개)

주택유형	자치구(군)	유의수준 5%	유의수준 1%
Type1	74	19	10
Type2	74	25	14
Type3	74	21	7
Type4	74	24	13
Type5	74	21	12

마. 기타 분류지표

- (층별 분류지표 선정) 경찰상태와 연관성이 높은 분류지표 선정
  - 고용조사 표본은 경찰표본과 별도표본을 결합한 형태이므로, 경찰조사의 표본설계 시 적용한 분류지표\*를 중심으로 시군의 특성을 반영\*\*하여 탐색
    - \* 15~29세비율, 3~40대비율, 60세이상비율, 자가비율, 농림어가비율, 1인가구비율, 30세이상대졸자비율, 대졸자비율, 주택유형, 서비스업취업자비율, 광제조업취업자비율, 농림어업취업자비율
- (상관분석) 인구 및 가구특성 관련 주요 변수 및 경찰 분류지표에 대하여 고용률 및 실업률과의 상관분석 수행
  - 12개 변수에 대하여 유의미한 상관관계가 있는 것으로 확인
  - 특히 30~40세 이상 비율, 65세 이상 비율 등은 다른 변수들에 비해 많은 층에서 유의미한 것으로 시현
- (회귀분석) 상관분석 결과 도출된 12개변수를 대상으로 층별 분류지표 선정을 위한 다중선형회귀분석\* 실시
  - \* 종속변수는 고용률임

- 다중선형회귀모형에 변수선택법(stepwise)\*을 적용하여 층별 분류지표 선정
  - \* 변수선택시 변수제거 0.1(SLS), 변수추가 0.2(SLE) 기준 적용
- (의사결정나무분석) 분류변수의 중요도 파악 및 구분점을 설정하기 위하여, 변수선택법으로 선정된 변수를 포함하여 층별 의사결정나무분석 수행

< 최종 분류지표 선정 결과 >

(단위 : 개)

		선정 결과		
		1순위	2순위	3순위
인구	15-29세 인구비율	-	-	-
	30-40대 인구비율	43	43	-
	50-60대 인구비율*	14	4*	10
	65세이상 인구비율	23	21	2
	30-40대 대졸이상 인구비율	6	1	3
	대졸이상 인구비율	-	-	-
	여성 인구비율	1	-	1
가구	가구원수	8	2	6
	유배우 비율	4	-	4
	농림어가 비율	5	1	3
	1인가구 비율	7	2	5
	자가 비율	-	-	-

\* 50-60대 인구비율은 14개 설계층에서 분류지표로 사용되었으며, 1순위로 사용된 설계층은 4곳임을 의미

### 3. 표본규모

- 특·광역시 내 자치구(군) 통계 작성을 위한 최소 표본규모 및 예산 등을 종합적으로 고려하여 3,120개 조사구(별도)로 결정
  - 경찰조사 표본조사구(738개 조사구\*)를 포함하여 특·광역시 고용조사 표본조사구는 3,858개 조사구임
    - \* 연동표본제 적용에 따라 2개 조사구에서 조사가 가능하므로 하반기 조사 시 경찰조사의 표본조사구는 820개 조사구임
  - '18년 상반기 고용조사와 분산이 동일하다는 가정 하에서 '19년 상반기 고용조사의 전국 실업자 상대표준오차(이하 RSE)는 1.6%로 예상

$$RSE_2 = \sqrt{\frac{n_1}{n_2}} \times RSE_1$$

여기서  $n_1$ 는 전년도 표본수,  $n_2$ 는 금년도 표본수이며,  $RSE_1$ 은 전년도 상대표준오차  $RSE_2$ 는 예상 상대표준오차임

#### 4. 표본배분

- (고려사항) 고용조사의 표본규모는 동시점에 조사되는 경찰조사와의 정합성 및 공표수준을 고려하여 결정될 필요
  - 기존 고용조사 표본(시군), 최근 고용조사 결과 및 시도 수준에서의 표본규모 등을 고려하여 표본규모 설정

< 시도별 조정된 표본규모 및 시군구수 >

(단위: 천명, 조사구, 개)

	만15세 이상 인구	표본규모				시군구
		'17 개편	자치구 확대	최소*	최대**	
특·광역시	19,565.2	2,402	3,857	-	-	74
서울	8,516.4	771	1,227	37	62	25
부산	2,980.5	404	724	35	61	16
대구	2,097.5	306	450	37	80	8
인천	2,452.4	307	542	35	75	10
광주	1,255.0	209	321	47	76	5
대전	1,292.4	221	317	52	81	5
울산	971.0	184	276	50	67	5
세종	190.7	100	100	-	-	1

\* 이전 표본규모 하한(35개)을 적용

\*\* 시도 내 경찰조사 표본은 시군구를 자유롭게 이동할 수 있어 최대 표본규모는 축소될 수 있음

- (목표정도) '18년 상반기 고용조사와 분산이 동일하다는 가정 하에서 '19년 상반기 고용조사의 특·광역시별 실업자 RSE는 3~8% 내외로 예상
  - 특·광역시 내 74개 구(군)의 실업자 RSE는 15~30% 수준으로 예상

< 시도별 표본 증감 및 예상 RSE >

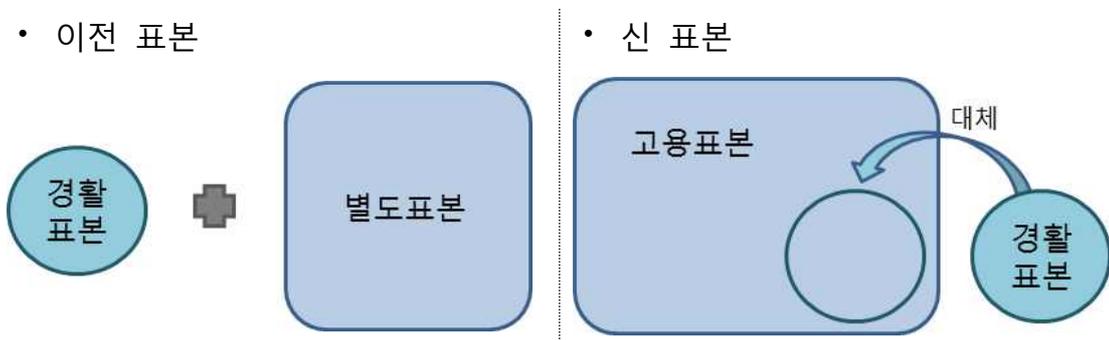
(단위: 조사구, %)

시도	2017년 개편				표본 확대					
	표본 (a)	경찰	별도 (b)	RSE (실업자)	표본 (c)	증감 (c-a)	경찰 (4월)	별도 (d)	증감 (d-b)	예상 RSE (실업자)
특광역시	2,406	738	1,668	-	3,858	1,452	738	3,120	1,452	-
서울	771	207	564	4.0	1,227	456	207	1,020	456	3.1
부산	404	108	296	5.7	724	320	108	616	320	4.2
대구	306	90	216	6.9	450	144	90	360	144	5.7
인천	307	99	208	6.5	543	236	99	444	236	4.9
광주	213	81	132	10.1	321	108	81	240	108	8.2
대전	221	81	140	9.2	317	96	81	236	96	7.7
울산	184	72	112	8.3	276	92	72	204	92	6.8

## 5. 표본추출

- 표본추출 방법 : 층화2단집락추출(Stratified Two-stage Cluster Sampling)
  - 표본조사구 : 표본조사구를 확률비례계통추출
    - 크기척도 : 조사구의 가구 수
    - 분류지표 : 행정구역(동·읍·면)과 설계층별 분류지표 활용
  - 표본가구
    - 추출된 표본조사구 내 시작가구를 단순임의추출
    - 시작가구를 포함한 연속된 20가구를 표본가구로 선정
  - 표본대체
    - 고용조사는 경찰조사와 별도 표본을 함께 이용함에 따라, 고용조사 표본 추출 후 설계층별 분류지표를 기준으로 일부 표본을 경찰표본으로 대체
      - 시도 수준에서 추출된 경찰표본과 자치구(군) 수준에서 추출된 별도표본을 단순히 통합할 경우, 자치구(군) 수준에서의 대표성에 문제 발생 가능
      - 이에 고용조사 표본을 자치구(군) 수준에서 추출한 후, 추출된 표본 중 일부를 경찰표본으로 대체\*
- \* 조사구 특성이 유사한 표본 간 대체가 이루어질 수 있도록 분류지표 기준으로 탐색

< 고용조사 표본대체 개념도 >



## 6. 가중값 작성 및 추정

### 가. 설계가중값

- 고용조사 표본 중 일부를 경할표본으로 대체하여 사용함에 따라, 경할표본에 대한 별도의 설계가중값 고려 불필요

$$\circ w_{hij}^0 = \frac{M_h}{n_h M_{hi}} \times \frac{M_{hi}}{m_{hi}} = \frac{M_h}{n_h \times 20} = w_{hijk}^0$$

$\uparrow$                        $\uparrow$   
 층별                      가구  
 표본조사구              추출률  
 추출률 역수              역수

- $w_{hij}^0$  :  $h$ 층의  $i$ 번째 표본조사구 내  $j$ 번째 표본가구의 설계가중값
- $h$  : 설계층(162개 시도 및 시군,  $h = 1, \dots, 162$ )
- $i$  : 표본조사구 ( $i = 1, 2, \dots, n_h$ )
- $j$  : 표본가구 ( $j = 1, 2, \dots, m_{hi}$ )
- $k$  : 표본가구 내 조사대상 가구원 ( $k = 1, \dots, l_{hij}$ )
- $M_h$  :  $h$ 층의 전체 가구수
- $M_{hi}$  :  $h$ 층  $i$ 번째 표본조사구의 전체 가구수
- $n_h$  :  $h$ 층의 표본조사구수
- $m_{hi}$  :  $h$ 층  $i$ 번째 표본조사구 내 표본가구수(20가구)

### 나. 사후가중값

- 설계층·성·연령대별로 사후 층화하여 현재(추계) 인구에 맞게 조정

$$\circ w_{hijk} = w_{hijk}^0 \times \frac{X_{h,sa}}{\hat{X}_{h,sa}}$$

- $h$  : 설계층( $h = 1, 2, \dots, 74$ )
- $X$  : 벤치마킹 모집단 추정값(현재 인구)
- $\hat{X}$  : 응답한 가구원의 설계가중값 합
- $sa$  : 사후층(성별 2개, 연령 5세별 11개 층)
- $w_{hijk}^b$  :  $h$ 층의  $i$ 번째 조사구 내  $j$ 번째 가구의  $k$ 번째 가구원의 설계가중값
- $w_{hijk}$  :  $h$ 층의  $i$ 번째 조사구 내  $j$ 번째 가구의  $k$ 번째 가구원의 사후가중값

## 다. 추정

### ◦ 총계 추정

$$\circ \hat{Y} = \sum_h \hat{Y}_h$$

$$\circ \hat{Y}_h = \sum_i \sum_j \sum_k w_{hijk} \cdot y_{hijk}$$

·  $w_{hijk}$  :  $h$ 층의  $i$ 번째 표본조사구 내  $j$ 번째 가구의  $k$ 번째 가구원의 사후가중값

### ◦ 표본오차 추정

$$\circ \hat{V}(\hat{Y}) = \sum_{h=1}^{74} \frac{n_h}{n_h - 1} (1 - f_h) \sum_{i=1}^{n_h} (y_{hi.} - \bar{y}_{h..})^2$$

·  $h$  : 설계층(경찰)

·  $N_h$  :  $h$ 층의 현재인구

·  $n_h$  :  $h$ 층의 표본 조사구수

·  $f_h$  : 조사구 추출률

$$\cdot y_{hi.} = \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}$$

$$\cdot \bar{y}_{h..} = \left( \sum_{i=1}^{n_h} y_{hi.} \right) / n_h$$

$$SE(\hat{Y}_h) = \sqrt{\widehat{Var}(\hat{Y}_h)}$$

$$RSE = \frac{SE(\hat{Y}_h)}{\hat{Y}_h} \times 100 (\%)$$