

2020 인구주택총조사 표본설계

2020. 9.

황현식, 김건수 (통계청 표본과)

 목 차

I. 개요 ¹

II. 지난주기 대비 주요 변동사항 ³

III. 표본 설계 ⁵

IV. 표본 추출 결과 ²⁰

V. 가중값 작성 ²³

VI. 결론 및 토의 ²⁵

I 개요

- 조사목적
 - 국가 주요 정책의 기초 자료로 활용
 - 각종 표본조사의 모집단 및 표본틀로 활용 등
- 조사대상
 - 조사기준 시점 현재 대한민국 영토 내에 상주하는 모든 내·외국인과 이들이 살고 있는 거처
 - * 해외 취업·취학 중인 사람, 외국군대의 병영 막사 등 제외
- 조사항목 : 전수 16개 항목 및 표본 55개 항목
- 기준시점 : 2020. 11. 1. 0시 현재

- 조사기간
 - 인터넷 및 전화 조사 : '20. 10. 15. ~ 10. 31.(17일간)
 - 방문 면접조사 : '20. 11. 1. ~ 11. 18.(18일간)
- 조사주기 : 5년
- 조사방법
 - 전수(등록센서스): 행정자료 수집
 - 표본(현장조사): 인터넷, 전화, 면접조사

- **표본설계의 주요 포인트**
 - 20%의 표본으로 시군구(250개) 읍면동(3,489개) 공표
 - 취약조사구(외국인, 1인가구, 고시원 등) 고려
 - 등록자료('18년 11월 기준 등록기반 가구부문 표본추출틀) 활용가능성 모색

II 지난주기 대비 주요 변동사항

1 조사구 규모 확대

- 30가구 내외 → 60가구 내외
 - 조사구 내 가구 수가 많을수록 과소조사구 발생률이 줄어드는 장점이 있음
 - 조사구 규모를 크게 하는 것이 대체 조사구를 추출하는 것보다 효율적
 - 30가구 기준 조사구 활용에 따른 우려 발생

* 소규모 조사구 활용으로 인한 영향(30가구)

- 추출단위의 변동성 확대(시간경과에 따른 가구수 감소 등)
- 조사구 내 높은 집락효과 발생으로 표본규모 확대 필요성 상승
- 다문화, 장애인 가구 등 발견률이 낮은 경우 조사대상 확보에 어려움 발생
- 조사구 내 부적격, 빈집, 불응 등에 따른 유효표본 부족
- 현 조사구 내 추출률 유지*에 어려움

* 일정 규모 미만의 조사구는 과소조사구로 표본에서 제외

2 표본 배분체계

- 읍면동 직접 배분 → 시군구, 읍면동 단계적 배분
 - 읍면동에 직접 표본을 배분하는 경우 층별 추출률 차이가 크고, 예상오차가 크게 발생
 - 시군구 내 읍면동수 및 읍면동별 인구규모 차이로 인한 표본규모 불균형으로 인해 작성수준별 예상오차가 커짐

3 표본 추출방법

- 계통추출 → 확률비례계통추출
 - 추출단위인 조사구 내 가구수에 차이(40~90가구)가 있으므로, 확률비례추출(PPS) 방법을 적용 → 가구수가 많으면 추출확률이 높아짐

Ⅲ 표본 설계

1 조사모집단

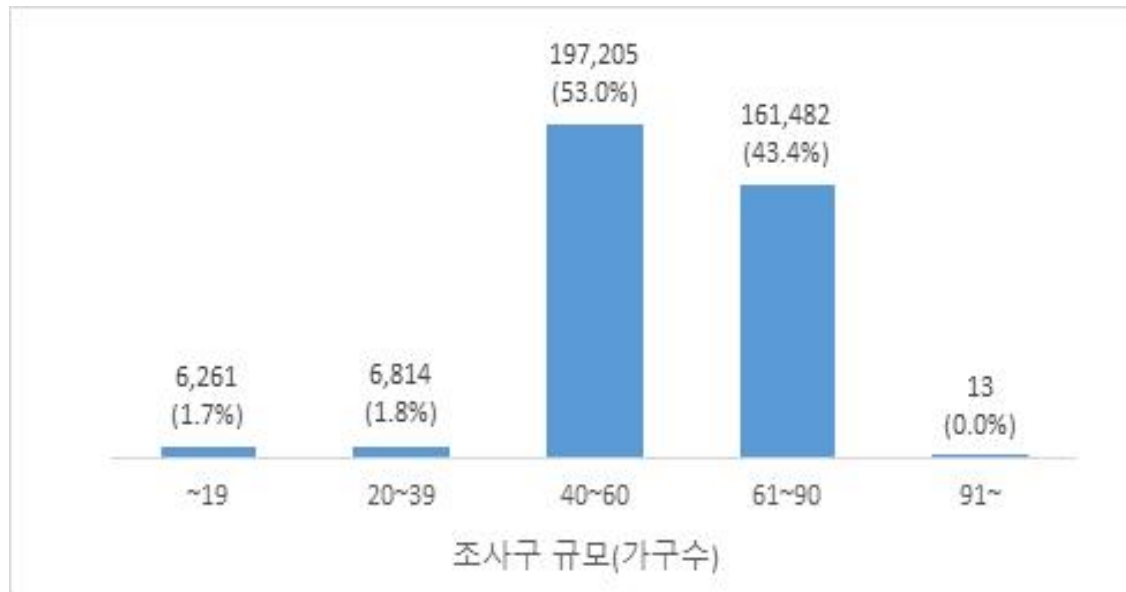
- '19년 가구주택기초조사 조사구에 거주하는 모든 가구, 가구원 및 이들이 거주하는 거처

< 표1 : 조사구 특성별 현황 >

조사구 특성	조사구수(개)	구성비(%)	비 고
전 체	381,705	100.0	
소 계	371,775	97.4	20% 표본조사
아파트(A)	190,504	49.9	
보통(1)	180,358	47.3	
섬(2)	913	0.2	
기숙·사회시설(3, 4)	9,930	2.6	전수조사

- 조사구 규모는 평균 60가구로 40~90가구가 전체의 96.5%를 차지

< 그림1 : 조사구 규모별 분포 >



* 아파트(A), 보통(1), 섬(2) 조사구 대상

2 표본추출틀

- '19년 가구주택기초조사 조사구 중 특성이 아파트(A), 보통(1), 섬(2) 조사구
 - (제외) 20가구 미만의 과소조사구 및 사용조사구* 일부
 * 인층 시범예행조사, 시험조사, 지역별고용조사 등

< 표2 > 표본추출틀의 조사구 포함률

(단위: 개, %)

조사모집단			표본추출틀					
전체	아파트	보통·섬	전체	아파트		보통·섬		
			포함률	포함률	포함률	포함률		
371,775	190,504	181,271	344,426	92.6	177,496	93.2	166,930	92.1

○ 표본추출틀 분포

- 읍면동 내 조사구수는 20~49개인 읍면동이 870개로 전체의 24.9%를 차지
- 경기도 부천시 신중동이 773개로 가장 크고,
서울시 강동구 둔촌1동, 경기도 파주시 진동면, 강원도 철원군 근북면이 1개로 가장 작음

< 표3 > 읍면동내 조사구수 분포

(단위: 개, %)

층	전체 조사구	읍면동 수	읍면동 내 조사구 수						
			~9	10~19	20~49	50~99	100~149	150~299	300~
전 국	344,426	3,489 (100.0)	25 (0.7)	478 (13.7)	870 (24.9)	691 (19.8)	606 (17.4)	730 (20.9)	89 (2.6)
서 울	64,041	425	2	3	12	65	145	190	8
부 산	23,435	205	2	4	22	74	47	53	3
대 구	15,537	141	0	1	16	60	36	25	3
인 천	18,866	154	1	9	26	16	50	52	0
광 주	9,410	96	0	4	26	29	20	16	1
대 전	9,832	79	0	1	7	23	29	16	3
울 산	7,540	56	0	1	3	15	15	20	2

< 표3 > 읍면동내 조사구수 분포(계속)

(단위: 개, %)

층	전체 조사구	읍면동 수	읍면동 내 조사구 수						
			~9	10~19	20~49	50~99	100~149	150~299	300~
세 종	2,091	19	0	1	5	5	3	3	2
경 기	83,961	546	4	16	60	110	122	197	37
강 원	10,668	188	3	28	91	35	17	14	0
충 북	11,270	153	0	36	47	31	22	12	5
충 남	15,097	207	0	33	90	35	21	22	6
전 북	12,614	243	4	84	85	32	18	18	2
전 남	12,793	297	4	89	140	38	15	9	2
경 북	19,330	332	2	87	141	45	19	35	3
경 남	23,439	305	3	78	88	65	23	39	9
제 주	4,502	43	0	3	11	13	4	9	3

○ '18년 기준 등록센서스 정보 활용

- '19년 가구주택기초조사 조사구 + '18년 기준 등록기반 추출틀 조사구
- 조사구 경계 정보(면, 중심점)를 이용하여 조사구 단위로 연계
- 분류지표 및 사용조사구 여부 확인에 활용

< 그림2 : 조사구 연계 예시 >



3 표본 규모

- 표본규모
 - 전체 조사구의 20%
 - 기본(18.5%) + 취약조사구*(1.5%)

* 아래 변수에 대해 각 읍면동별 순위를 부여 후 배정

①1인 가구수, ②다가구 주택수, ③오피스텔수, ④외국인수, ⑤고시원·고시텔수

< 표4 > 시도별 표본조사구 규모 현황

(단위: 개)

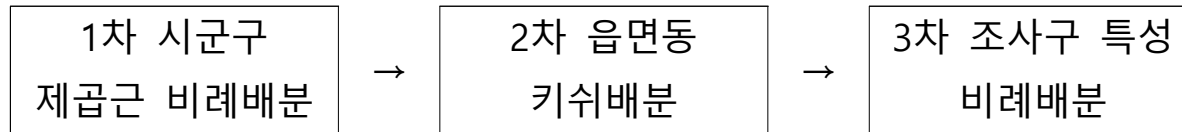
지 역	최종(20%)			기본(18.5%)			취약(1.5%)		
	전체	아파트	보통·섬	전체	아파트	보통·섬	전체	아파트	보통·섬
전 국	74,355	31,113	43,242	68,704	30,102	38,602	5,651	1,011	4,640
서 울	11,085	4,347	6,738	9,471	4,163	5,308	1,614	184	1,430
부 산	4,852	2,631	2,221	4,700	2,605	2,095	152	26	126
대 구	2,885	1,520	1,365	2,723	1,496	1,227	162	24	138
인 천	3,753	1,793	1,960	3,382	1,715	1,667	371	78	293
광 주	1,784	1,088	696	1,756	1,082	674	28	6	22
대 전	1,709	885	824	1,617	871	746	92	14	78
울 산	1,623	950	673	1,564	945	619	59	5	54
세 종	378	245	133	354	239	115	24	6	18

< 표4 > 시도별 표본조사구 규모 현황(계속)

(단위: 개)

지 역	최종(20%)			기본(18.5%)			취약(1.5%)		
	전체	아파트	보통·섬	전체	아파트	보통·섬	전체	아파트	보통·섬
경 기	15,437	7,508	7,929	13,320	7,156	6,164	2,117	352	1,765
강 원	3,200	1,186	2,014	3,168	1,160	2,008	32	26	6
충 북	2,876	1,128	1,748	2,822	1,105	1,717	54	23	31
충 남	3,902	1,403	2,499	3,662	1,335	2,327	240	68	172
전 북	3,778	1,193	2,585	3,715	1,169	2,546	63	24	39
전 남	4,668	1,084	3,584	4,556	1,032	3,524	112	52	60
경 북	5,738	1,631	4,107	5,562	1,590	3,972	176	41	135
경 남	5,950	2,300	3,650	5,634	2,235	3,399	316	65	251
제 주	737	221	516	698	204	494	39	17	22

4 표본 배분



- (1차) 시군구 배분
 - 제공근 비례배분(조사구수), 2015년 표본오차를 감안하여 일부 조정
 - * 예상 RSE 특광역시 20.0%, 도내 시지역 20.0%, 군지역 25.0% 이하
- (2차) 읍면동 배분
 - 키쉬배분(조사구수), 2015년 표본오차를 감안하여 일부 조정
 - 최소 9개 표본조사구 배분
- (3차) 조사구 특성별 배분
 - 조사구 특성(①아파트 ②보통·섬)별 비례배분
 - 최소 2개 표본조사구 배분

☞ (참고) **제공근비례배분**

- ✓ 시뮬레이션 결과 제공근 배분 방식으로 결정('19.9.)
- ✓ 제공근은 조사구수 활용

※ 가구수 제공근에 비해 규모가 작은 시군구의 추출률을 높임

[**역(Power) 배분**]

$$n_h = n \times \frac{N_h^P}{\sum_{h=1}^H N_h^P}$$

[**제공근(P=0.5) 배분**]

$$n_h = n \times \frac{\sqrt{N_h}}{\sum_{h=1}^H \sqrt{N_h}}$$

(n : 표본조사구수, h : 1~ H 층, N : 모집단 조사구수)

☞ (참고) kish 가중최적배분

✓ 조사구수에 적용¹⁾

※ 제곱근배분: 면지역 특성이 많이 반영된 표본 배정, 시군구 통계의 질 저하

※ 비례배분: 규모가 작은 면지역에 과소표본 발생 우려

※ kish 배분: 일정규모 이하와 이상의 경우 추출률이 더 높고, 중간규모의 경우 더 낮은 특징을 보임

[kish 가중최적배분]

$$n_h = n \times \frac{\sqrt{I \times W_h^2 + J \times H^{-2}}}{\sum_{h=1}^H \sqrt{I \times W_h^2 + J \times H^{-2}}}$$

n : 표본조사구수, h : 1~ H 층, N : 모집단 조사구수

$W_h = \frac{N_h}{N}$, H : 읍면동수, $I = \frac{I_1}{I_2}$ 는 층(읍면동) 단위 추정 대비 전국 수준 추정의 상대적 중요도로 $I=0$ 이면

균등배분과 동일하고, $I \rightarrow \infty$ 이면 비례배분과 동일. 보통 $I=1$ 로 설정하지만 여기서는 $I=0.9$, $J=0.1$ 로 설정함

1) LESLIE KISH. Multipurpose Sample Design. Survey Methodology, June 1988, Vol. 14, No. 1, pp. 19-32, Statistics Canada

5] 층화 및 분류지표

- 층화: 읍면동 × 조사구 특성(A, 1*), 층수(6,143개)
 - * 조사구 특성 섬(2)은 보통(1)과 통합
- 분류지표
 - 연계된 **2018년 기준 등록기반 가구부문 표본추출틀**의 정보를 활용
 - 1인 가구와 고령자의 증가 추세를 반영하여 1인 가구 비율, 65세이상 인구 비율 등
 - 층별 중앙값의 이상/미만으로 구간화

< 표5 > 분류지표 목록

변수 분류	분류지표	정보 출처
조사구 특성	아파트(A), 보통·섬(1, 2)	2019년 가구주택기초조사
단독 비율	설계층별 단독주택 가구 비율로 구간화	
조사구 번호	조사구 번호	
1인 가구 비율	설계층별 1인 가구 비율로 구간화	2018년 기준 등록기반 가구부문 표본추출틀
65세 이상 인구 비율	설계층별 65세 이상 인구 비율로 구간화	

< 표6 > 분류지표 정렬 순서

조사구 특성	분류지표 순서
아파트	① 1인 가구 비율 ② 65세 이상 비율 ③ 조사구 번호
보통·섬	① 조사구 특성(1,2) ② 단독 비율 ③ 1인 가구 비율 ④ 65세 이상 비율 ⑤ 조사구 번호

6 표본 추출

- 추출방법: 확률비례계통추출
 - 층별 분류지표로 정렬 후, 표본규모 만큼 확률비례계통추출
 - 크기 변수(MOS): $\log(\text{가구수})^*$
 - * 추출률이 높아서(추출간격이 좁아서) 2번 이상 추출되는 조사구 방지

IV 표본 추출 결과

< 표7 > 시도별 표본조사구 추출 현황

(단위: 개, %)

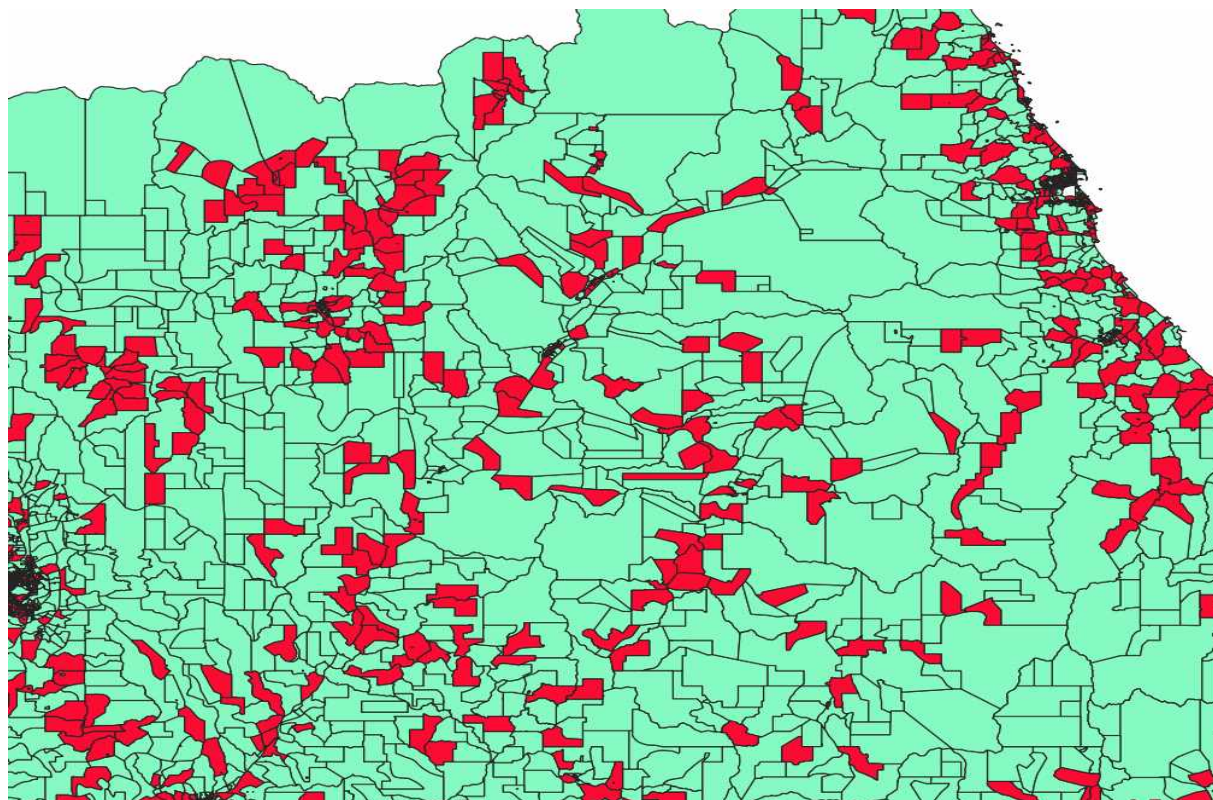
지 역	조사모집단			표 본					
	전체	아파트	보통·섬	전체	추출률	아파트	추출률	보통·섬	추출률
전 국	371,775	190,504	181,271	74,355	20.0	31,113	16.3	43,242	23.9
서 울	68,159	29,803	38,356	11,085	16.3	4,347	14.6	6,738	17.6
부 산	25,599	14,822	10,777	4,852	19.0	2,631	17.8	2,221	20.6
대 구	16,765	9,741	7,024	2,885	17.2	1,520	15.6	1,365	19.4
인 천	19,996	10,900	9,096	3,753	18.8	1,793	16.4	1,960	21.5
광 주	10,124	6,798	3,326	1,784	17.6	1,088	16.0	696	20.9
대 전	10,674	5,884	4,790	1,709	16.0	885	15.0	824	17.2
울 산	8,348	5,133	3,215	1,623	19.4	950	18.5	673	20.9
세 종	2,364	1,807	557	378	16.0	245	13.6	133	23.9

< 표7 > 시도별 표본조사구 추출 현황(계속)

(단위: 개, %)

지 역	조사모집단			표 본					
	전체	아파트	보통·섬	전체	추출률	아파트	추출률	보통·섬	추출률
경 기	87,840	49,809	38,031	15,437	17.6	7,508	15.1	7,929	20.8
강 원	12,289	5,774	6,515	3,200	26.0	1,186	20.5	2,014	30.9
충 북	12,352	6,070	6,282	2,876	23.3	1,128	18.6	1,748	27.8
충 남	16,451	7,575	8,876	3,902	23.7	1,403	18.5	2,499	28.2
전 북	13,970	6,605	7,365	3,778	27.0	1,193	18.1	2,585	35.1
전 남	14,585	5,666	8,919	4,668	32.0	1,084	19.1	3,584	40.2
경 북	21,683	9,074	12,609	5,738	26.5	1,631	18.0	4,107	32.6
경 남	25,497	13,451	12,046	5,950	23.3	2,300	17.1	3,650	30.3
제 주	5,079	1,592	3,487	737	14.5	221	13.9	516	14.8

< 그림3 : 표본조사구 지역 분포 예시 >



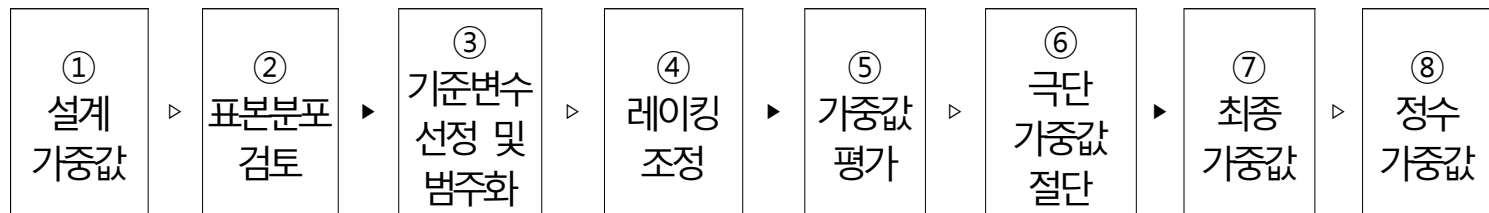
* 붉은색이 표본조사구, 강원도 일부 지역

V 가중값 작성

1 가중값 작성 대상 및 절차

- 작성대상
 - (인구) 내국인, 외국인
 - (가구) 일반·집단가구, 외국인가구
 - (거처) 사는 집, 빈집

○ 작성 절차



2 주요 절차

- 기준변수 선정
 - 가중값을 적용하기 위해 모집단을 가장 대표할 수 있는 핵심변수를 기준변수로 적용
- 레이킹 조정
 - 모집단의 알려진 주변분포에 맞게 반복적으로 추정값 조정
- 가중값 평가
 - 모집단과 표본의 분포 차이에 의해 극단가중값 발생
- 가중값 정수화
 - 이용자 편의성 및 이해도 제고를 위해 정수로 제공
 - 정수와 소수로 분리하여 정렬 후 소수 크기에 따라 정수 가중값 변형

ID	최종w	정수	소수	추가	정수w
1	2.3	2	0.3	-	2
2	3.3	3	0.3	-	3
3	4.4	4	0.4	1	5
합계	10.0	9	1.0		10

VI 결론 및 토의

1 '20년 인구주택총조사의 표본설계 방법 정리

- 조사구 규모(60가구), 표본배분(제공근비례, kish), 표본추출(확률비례계통)
- 분류지표(1인가구비율, 65세이상인구비율 등)

2 가능한 모든 정보의 과학적 활용

- 2019년 가구주택기초조사
- 2018년 기준 등록기반 가구부문 표본추출틀
- 조사구역도 등

☞ 토의 : 잘 추출되었는지 확인할 지표개발, 더 고려할 사항은 없는가?