
2020년 기준 「중장년층통계」 소득·부채항목
작성을 위한 표본설계 결과

2021. 10.



통계청
조사관리국 표본과

목 차

I. 증장년층통계 개요	1
1. 개요	1
2. 표본설계 의뢰내용	2
II. 표본 설계	3
1. 모집단 분석	3
2. 층화 및 표본배분	4
3. 표본 추출	7
III. 추정	9
1. 설계 가중값	9
2. 사후 가중값	9
3. 추정식	10
[붙임] 설계층별 추출률	11

1. 개요

- (작성목적) 중장년층의 실태를 파악할 수 있는 통계를 개발하여 은퇴 후 재취업 및 노후생활안정을 위한 정책수립의 기초자료 제공
- (작성유형) 가공통계(행정통계)
- (모집단 및 작성대상) 매년 11월 1일 기준 국내 거주하는 만40~64세 연령의 내국인과 그 내국인을 포함하는 일반가구
 - * 외국인가구, 집단가구, 집단시설가구는 제외
- (작성주기 및 기준시점) 연간 / 매년 11월 1일 0시 기준
- (작성단위) 개인, 가구
- (활용 행정자료) 인구·가구·주택DB 자료, 주택소유통계DB 자료, 일자리 행정통계DB 자료, 건강보험 직장가입자DB 자료, 공적연금 자료, 퇴직연금통계DB 자료, 소득자료^(국세청), 부채자료^(신용정보원) 등 14종
- (작성 항목) 성별, 연령 등 기본현황을 포함한 30개 항목 작성

	작성 항목
(1) 기본현황	① 거주지 ② 가구주와의 관계 ③ 성별 ④ 연령 ⑤ 세대구성유형 ⑥ 가구원수 ⑦ 거주유형 ⑧ 주거면적 ⑨ 가구주 성별 ⑩ 가구주 연령 ⑪ 자녀 연령
(2) 경제활동	① 종사상지위 ② 근속기간 ③ 산업대분류 ④ 종사자규모 ⑤ 월 평균 임금 ⑥ 전년 종사상 지위 ⑦ 전년 근속기간 ⑧ 전년 산업대분류 ⑨ 전년 종사자규모 ⑩ 전년 월 평균 임금 ⑪ 개인기업체 창업
(3) 주택소유	① 개인 주택 소유물건수 ② 개인 주택 소유자산가액 ③ 가구 주택 소유물건수 ④ 가구 주택 소유자산가액
(4) 소득 및 금융권 대출잔액	① 소득금액 ② 금융권 대출잔액
(5) 연금가입	① 공적연금 가입 여부 ② 퇴직연금 가입여부

- (모집단 및 표본추출틀) '20년 11월 1일 기준 국내 거주하는 40~64세 연령의 내국인 (20,086,445명)

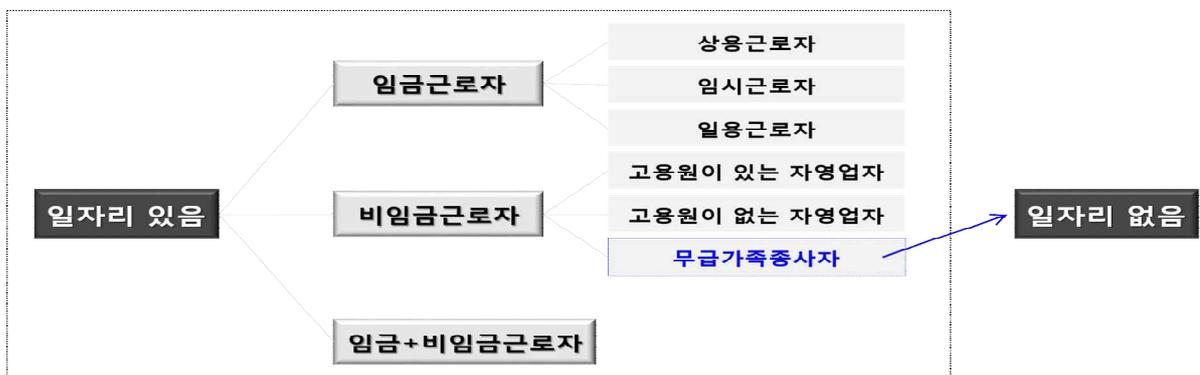
2. 표본설계 의뢰 내용

- 본 통계는 행정자료를 활용한 전수통계이나, 소득과 부채항목은 필요한 표본에 대해서만 자료입수가 가능하여 표본추출 필요
- 공표단위별 소득·부채항목 통계생산이 가능한 표본설계
 - 시도(17)×성(2)×연령(5), 고용형태(4), 주택소유(2) × (소득·부채)

- (시 도) 서울/부산/.../경기/강원/.../제주
- (성 별) 남자/여자
- (연 령) 40-44세/45-49세/50-54세/55-59세/60-64세
- (고용형태) 등록취업자(임금근로자/비임금근로자/임금·비임금근로 병행)/미취업(미등록)자
- (주택소유) 주택소유/주택미소유
- (소득구간) 소득있음(1천미만/3천미만/5천미만/7천미만/1억미만/1억이상)/소득없음
- (부채구간) 부채있음(1천미만/3천미만/5천미만/7천미만/1억미만/2억미만/3억미만/3억이상)/부채없음

- 표본규모 : 약 20만 명(모집단의 약 1%)
- 일자리DB 및 주택소유통계DB 구축시점과의 차이로 인해 표본추출 단계에서 고용형태와 주택소유여부 항목은 전년도 정보 활용 가능
 - 추후 해당DB 구축이 완료되면 사후가중치 산출 요청

※ 취업자 중 무급가족종사자는 비임금근로자에 해당하나, 행정자료를 통해 파악이 어려워 「중장년층통계」에서는 '일자리없음'으로 분류



※ 사회보험에 가입하지 않았거나 근로소득 신고가 되지 않은 근로자, 무등록사업체 종사자, 무급가족종사자, 농림어업생산활동경영인 등 행정자료로 파악이 곤란한 대상은 등록취업자에서 제외되어 있음

1. 모집단 분석

- (모집단) '20. 11. 1. 기준, 국내 거주 40~64세 내국인 20,086,445명
- (고용형태) '19년 기준 비임금(11.9%)과 임금+비임금(2.2%) 비중이 낮음
 - (시도) 세종·울산의 임금근로자 비중이 높고, 전남·경북이 가장 낮음
 - (성, 연령) 연령이 낮을수록, 남성이 여성보다 임금근로자 비중이 높음
 - (주택소유) 주택을 소유한 경우 임금근로자 비중이 높음

<표1> 모집단 분포

구분	중장년층 인구 (1,000명)	비중 (%)				
		시도/특성별	고용형태별			
			임금	비임금	임금+비임금	일자리 없음
전국	20,086	100.0	50.6	11.9	2.2	35.3
서울	3,565	17.7	50.5	11.6	2.0	36.0
부산	1,330	6.6	48.2	12.4	1.8	37.6
대구	976	4.9	47.8	13.1	2.2	36.9
인천	1,180	5.9	51.4	11.7	1.9	34.9
광주	572	2.8	52.5	11.2	2.9	33.4
대전	573	2.9	52.7	11.4	2.2	33.7
울산	475	2.4	53.7	10.4	1.6	34.4
세종	126	0.6	59.4	9.6	2.0	29.0
경기	5,244	26.1	52.5	11.9	2.1	33.5
강원	609	3.0	49.8	12.2	2.6	35.4
충북	625	3.1	52.0	11.8	2.4	33.9
충남	800	4.0	50.6	11.9	2.4	35.0
전북	688	3.4	49.6	11.5	2.9	35.9
전남	685	3.4	47.6	11.3	2.6	38.5
경북	1,035	5.2	47.0	12.0	2.3	38.7
경남	1,346	6.7	49.1	12.4	2.1	36.3
제주	259	1.3	48.6	13.8	2.9	34.8
남자	10,084	50.2	58.4	13.7	2.7	25.1
여자	10,003	49.8	42.8	10.0	1.6	45.6
40-44세	3,758	18.7	58.0	9.5	2.1	30.4
45-49세	4,195	20.9	55.9	11.2	2.6	30.3
50-54세	4,246	21.1	53.1	12.4	2.5	32.0
55-59세	4,092	20.4	47.6	13.3	2.1	37.1
60-64세	3,795	18.9	38.1	12.8	1.5	47.6
주택없음	11,623	57.9	46.6	10.2	1.8	41.3
주택있음	8,464	42.1	56.2	14.1	2.6	27.0

* 고용형태 및 주택소유여부는 '19년 기준

2. 총화 및 표본배분

- (표본규모) 20만 명
 - (총화) 공표수준을 고려해서 시도(17)×고용형태(4)를 총화변수로 선정(68개 층)
 - (표본배분) 상세 통계별 분석 공표가 가능하도록 시도×고용형태별 제공근비례 배분
 - 표본설계 당시 '20년 고용형태 분포에 대한 정보가 없어 전년대비 변동률*이 '19년과 동일함을 가정하고 시도별 고용형태 표본규모 산출
 - 목표 표본규모 : '19년 기준 고용형태 제공근비례 배분<표2>
 - 시도별 고용형태 변동률 : '18~'19년 고용형태 변동률<붙임>
- * '20년 고용형태가 있는 일자리행정통계DB가 11월 중 확정되므로 12월에 자료 입수 후 무응답 보정 및 사후 가중값 작성시 반영 예정

<표2> 1차 배분 결과(비중)_층별 목표 표본규모

구분	시도별 비중 (%)	고용형태별 비중 (%)				
		합계	임금	비임금	임금+비임금	일자리 없음
전국	100.0	100.0	39.5	19.1	8.3	33.1
서울	11.2	100.0	39.7	19.0	7.8	33.5
부산	6.9	100.0	38.7	19.6	7.5	34.2
대구	5.9	100.0	38.2	20.0	8.2	33.6
인천	6.4	100.0	40.1	19.1	7.8	33.0
광주	4.5	100.0	40.1	18.5	9.4	32.0
대전	4.5	100.0	40.5	18.9	8.2	32.4
울산	4.0	100.0	41.5	18.2	7.1	33.2
세종	2.1	100.0	43.8	17.6	8.0	30.6
경기	13.6	100.0	40.4	19.2	8.1	32.3
강원	4.7	100.0	38.9	19.3	8.9	32.8
충북	4.7	100.0	40.1	19.1	8.5	32.3
충남	5.3	100.0	39.5	19.1	8.6	32.8
전북	5.0	100.0	38.8	18.7	9.4	33.0
전남	5.0	100.0	38.2	18.6	9.0	34.3
경북	6.1	100.0	38.0	19.2	8.4	34.5
경남	6.9	100.0	38.9	19.6	8.1	33.5
제주	3.1	100.0	38.1	20.3	9.3	32.3

- <표3>의 '18~'19년 고용형태 변동현황을 보면 임금근로자(87.8%), 비임금근로자(82.1%), 일자리없음(80.8%), 임금+비임금(59.6%) 순으로 전년도 고용형태를 유지하는 것으로 나타남

<표3> '18년~'19년 고용형태 변동 현황

구분		'18년 고용형태 변동률(%)			
		임금	비임금	임금+비임금	일자리없음
'19년 고용형태	임금	87.8	3.2	21.0	15.1
	비임금	0.8	82.1	14.6	3.8
	임금+비임금	0.9	2.9	59.6	0.3
	일자리없음	10.5	11.8	4.8	80.8
합계		100.0	100.0	100.0	100.0

* '20년 모집단 기준

- 고용형태 변동에 따라 전년 기준으로 표본추출 시, 모집단 비중이 낮은 “임금+비임금” 형태의 표본수 확보가 어려우므로 “임금+비임금” 형태에 대한 표본 배분 조정이 필요

※ '18~'19년 고용형태 변동을 이용한 표본규모 조정(붙임1 참고)

- 층화: 시도 × 고용형태
- 시도별 고용형태 표본 변동률*을 이용하여 '19년 기준으로 배분된 층별 표본규모 조정

* '19~'20년 고용형태 변동률은 '18~'19년 변동률과 동일하다고 가정

시도별 고용형태 표본비율 $X_s = A_s^{-1}B_s$, 표본수 $n_{s,i} = n_s \times x_{s,i}$

$$X_s = \begin{pmatrix} x_{s,1} \\ x_{s,2} \\ x_{s,3} \\ x_{s,4} \end{pmatrix} \quad A_s = \begin{pmatrix} a_{s,11} & a_{s,21} & a_{s,31} & a_{s,41} \\ a_{s,12} & a_{s,22} & a_{s,32} & a_{s,42} \\ a_{s,13} & a_{s,23} & a_{s,33} & a_{s,43} \\ a_{s,14} & a_{s,24} & a_{s,34} & a_{s,44} \end{pmatrix} \quad B_s = \begin{pmatrix} b_{s,1} \\ b_{s,2} \\ b_{s,3} \\ b_{s,4} \end{pmatrix}$$

- s : 시도
- i : $t-1$ 년도 고용형태
- $x_{s,i}$: s 시도 $t-1$ 년도 i 고용형태 표본비율
- $a_{s,ij}$: s 시도에서 $t-1$ 년도 i 고용형태에서 t 년도 j 고용형태로의 변동률 ($\sum_{j=1}^4 a_{ij} = 1$)
- n_s : s 시도 표본수 ($\sum_{i=1}^4 n_{s,i} = n_s$)
- j : t 년도 고용형태
- $b_{s,j}$: s 시도 t 년도 j 고용형태 표본비율

- (표본확정) '18~'19년 고용형태 변동률을 고려하면 모집단 비중이 낮은 “임금+비임금” 형태의 표본수가 약 46.7% 추가 배분됨

<표4> 2차 배분_층별 표본배분

구분	합계(명)	고용형태(명)			
		임금	비임금	임금+비임금	일자리없음
전국	200,000	71,606	38,649	24,294	65,451
서울	22,414	8,098	4,288	2,638	7,390
부산	13,721	4,886	2,689	1,578	4,568
대구	11,849	4,182	2,386	1,365	3,916
인천	12,890	4,664	2,422	1,563	4,241
광주	9,065	3,226	1,715	1,192	2,932
대전	8,994	3,259	1,724	1,089	2,922
울산	8,072	3,001	1,498	913	2,660
세종	4,145	1,617	734	507	1,287
경기	27,239	9,866	5,208	3,375	8,790
강원	9,375	3,305	1,856	1,155	3,059
충북	9,434	3,424	1,820	1,159	3,031
충남	10,695	3,817	2,060	1,335	3,483
전북	9,975	3,503	1,926	1,273	3,273
전남	9,927	3,436	1,896	1,235	3,360
경북	12,182	4,296	2,410	1,431	4,045
경남	13,856	4,907	2,751	1,650	4,548
제주	6,167	2,119	1,266	836	1,946

구분		합계	임금	비임금	임금+비임금	일자리없음
목표 표본수(A)	비중(%)	100.0	39.5	19.1	8.3	33.1
배분 표본수(B)	비중(%)	100.0	35.8	19.3	12.1	32.7
B/A		1.000	0.906	1.010	1.467	0.989

3. 표본 추출

1) 분류지표

- (분류지표) 층화변수를 제외하고 기본 공표단위인 성·연령과 소득 및 부채에 미치는 영향이 큰 거처유형, 주거용연면적, 가구원수를 이용하여 분류지표 구성

- 기타 거처유형의 경우 면적정보가 없어 단독주택 및 공동주택(아파트·연립·다세대)에 한해 면적 구분

분류 지표	범 주
① 성·연령	성(남 / 여) × 연령(40~44 / 45~49 / 50~54 / 55~59 / 60~64세)
② 거처유형· 주거용연면적	단독주택(60㎡ 이하 / 60㎡ - 165㎡ / 165㎡ 초과)
	아파트(60㎡ 이하 / 60㎡ - 100㎡ / 100㎡ 초과)
	연립·다세대(60㎡ 이하 / 60㎡ 초과)
	기타·미상
③ 가구원수	1인 / 2인 / 3인 / 4인 / 5인이상·미상

- (거처유형·연면적 병합) 층·거처유형·면적별로 예상되는 표본수가 작은 경우 인접 연면적 또는 거처유형 범주와 병합함

※ ① '광주, 임금+비임금, 연립·다세대 주택'의 경우, 모든 연면적을 병합하고,
 ② 세종특별자치시의 경우, 단독주택과 연립·다세대 주택을 병합한 후, 165㎡ 전후로 구별

- (분류지표의 유의성) 순위 로짓모형 분석 결과, 분류지표인 성·연령, 거처유형·주거용연면적, 가구원수는 소득, 부채에 대해 유의미한 차이를 보임

<표5> 순위 로짓모형 적합 결과

종속변수	독립변수	자유도	LR Statistic	Pr($\chi^2 > LR$)
소득	성·연령	9	2,637.2	<.0001
	거처유형·주거용연면적	11	2,845.2	<.0001
	가구원수	4	135.6	<.0001
부채	성·연령	9	1,556.3	<.0001
	거처유형·주거용연면적	11	1,267.6	<.0001
	가구원수	4	31.7	<.0001

* '19년 기준 결과 분석 / 소득, 부채 범주는 KOSIS 통계표 공표기준을 따름

2) 표본 추출

- (총화계통추출) 층별로 분류지표 순으로 정렬한 후 층별 표본수만큼 계통추출
- (추출결과) 기타 주요 공표단위인 성, 연령, 주택소유여부에 대해서도 모집단 분포와 차이가 크지 않음

<표6> 표본 분포

(단위 : 명, %, %p)

구분		모집단	비율(A)	표본	비율(B)	B-A
전국		20,086,445	100.0	200,000	100.0	0.0
시도	서울	3,564,564	17.7	22,414	11.2	-6.5
	부산	1,329,755	6.6	13,721	6.9	0.2
	대구	975,805	4.9	11,849	5.9	1.1
	인천	1,179,990	5.9	12,890	6.4	0.6
	광주	572,104	2.8	9,065	4.5	1.7
	대전	573,110	2.9	8,994	4.5	1.6
	울산	474,634	2.4	8,072	4.0	1.7
	세종	126,027	0.6	4,145	2.1	1.4
	경기	5,244,066	26.1	27,239	13.6	-12.5
	강원	608,745	3.0	9,375	4.7	1.7
	충북	624,820	3.1	9,434	4.7	1.6
	충남	799,724	4.0	10,695	5.3	1.4
	전북	687,696	3.4	9,975	5.0	1.6
	전남	685,378	3.4	9,927	5.0	1.6
	경북	1,035,466	5.2	12,182	6.1	0.9
	경남	1,345,519	6.7	13,856	6.9	0.2
제주	259,042	1.3	6,167	3.1	1.8	
고용형태	임금	10,172,930	50.6	71,606	35.8	-14.8
	비임금	2,384,479	11.9	38,649	19.3	7.5
	임금+비임금	435,643	2.2	24,294	12.1	10.0
	일자리없음	7,093,393	35.3	65,451	32.7	-2.6
성	남자	10,083,734	50.2	102,823	51.4	1.2
	여자	10,002,711	49.8	97,177	48.6	-1.2
연령	40-44세	3,758,298	18.7	35,795	17.9	-0.8
	45-49세	4,195,327	20.9	41,708	20.9	0.0
	50-54세	4,245,683	21.1	42,842	21.4	0.3
	55-59세	4,091,920	20.4	41,371	20.7	0.3
	60-64세	3,795,217	18.9	38,284	19.1	0.2
주택 소유	없음	11,622,822	57.9	113,048	56.5	-1.3
	있음	8,463,623	42.1	86,952	43.5	1.3

* 고용형태 및 주택소유는 '19년 기준

1. 설계 가중값

- 층별 추출률(대표성)이 다르기 때문에 자료 분석 및 통계작성을 위해서는 반드시 가중값 적용이 필요
- 설계층 내 표본가구들의 설계가중값은 동일

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

- w_{hi} : h 층 i 의 설계가중값(설계층 내 동일)
- h : 설계층 (시도*고용형태, 68개 층)
- i : 중장년층 개인
- N_h : h 층의 모집단 중장년 수
- n_h : h 층의 표본 중장년 수

<표7> 설계가중값 분포

표본수	최소값	중앙값	평균	최대값	표준편차	합 계
200,000	5.0	93.1	100.4	279.0	66.0	20,086,445

2. 사후 가중값

- (사후 보정) 표본추출 시 모집단 정보가 확정되기 전이므로 사후 모집단 확정자료로 사후보정 필요
- '20년 고용형태가 있는 일자리행정통계DB가 11월중 확정이므로 입수 후 무응답보정 및 사후 가중값 작성 예정

3. 추정식

○ 총계 추정

$$\hat{Y}_D = \sum_{h=1}^H \sum_{i=1}^{n_h} y_{hi} w_{hi} I[hi \in D]$$

- h : 설계층
- i : 표본(중장년)
- D : 관심영역(도메인)
- y : 관심변수
- w : 가중값
- n : 표본수
- N : 모집단수
- I : $\begin{cases} 1 & \text{if } hi \in D \\ 0 & \text{otherwise} \end{cases}$

○ 분산 추정

$$\hat{V}(\hat{Y}_D) = \frac{n_h(1-f_h)}{n_h-1} \sum_i^{n_h} \left\{ w_{hi} y_{hi} I[hi \in D] - \left(\sum_i^{n_h} w_{hi} y_{hi} I[hi \in D] \right) / n_h \right\}^2$$

$$f_h = n_h / N_h$$

$$SE(\hat{Y}_D) = \sqrt{\hat{V}(\hat{Y}_D)} = SE(\hat{Y}_D)$$

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

[붙임] 설계층별 추출률

	모집단 수 (1,000명) [P]				표본수 (명) [S]				추출률 (%) [f=S/P]			
	임금	비임금	임금+비임금	일자리 없음	임금	비임금	임금+비임금	일자리 없음	임금	비임금	임금+비임금	일자리 없음
서울	1,800	412	70	1,283	8,098	4,288	2,638	7,390	0.45	1.04	3.79	0.58
부산	640	165	24	500	4,886	2,689	1,578	4,568	0.76	1.63	6.50	0.91
대구	466	128	21	360	4,182	2,386	1,365	3,916	0.90	1.86	6.40	1.09
인천	607	138	23	412	4,664	2,422	1,563	4,241	0.77	1.76	6.83	1.03
광주	300	64	16	191	3,226	1,715	1,192	2,932	1.07	2.67	7.23	1.53
대전	302	65	12	193	3,259	1,724	1,089	2,922	1.08	2.63	8.78	1.51
울산	255	49	8	163	3,001	1,498	913	2,660	1.18	3.04	12.16	1.63
세종	75	12	3	36	1,617	734	507	1,287	2.16	6.05	20.09	3.53
경기	2,753	624	111	1,756	9,866	5,208	3,375	8,790	0.36	0.83	3.03	0.50
강원	303	75	16	215	3,305	1,856	1,155	3,059	1.09	2.49	7.24	1.42
충북	325	74	15	212	3,424	1,820	1,159	3,031	1.05	2.47	7.83	1.43
충남	405	95	19	280	3,817	2,060	1,335	3,483	0.94	2.16	6.99	1.24
전북	341	79	20	247	3,503	1,926	1,273	3,273	1.03	2.43	6.38	1.32
전남	326	77	18	264	3,436	1,896	1,235	3,360	1.05	2.45	6.87	1.27
경북	487	124	24	401	4,296	2,410	1,431	4,045	0.88	1.95	6.02	1.01
경남	661	167	28	489	4,907	2,751	1,650	4,548	0.74	1.64	5.83	0.93
제주	126	36	7	90	2,119	1,266	836	1,946	1.68	3.55	11.27	2.16