
「2022년 임금근로일자리 소득」 작성을 위한 표본설계 결과

■ 목 차 ■

- I. 통계 개요 1
- II. 표본설계 요청내용 2
- III. 표본설계 3
- IV. 추정 5

2023. 12.

조사관리국 표본과

I | 임금근로 일자리별 소득(일자리행정통계(소득편)) 개요

작성 목적

- 일자리 관련 정책 수립 및 일자리 선택에 유용한 기초자료 제공

모집단

- 기준연도 12월에 임금근로일자리에서 하루 이상 일한 근로자로서, 사회보험가입·근로소득신고 등으로 행정당국에 의해 파악된 자

작성 단위

- 일자리(기업체별 근로자)

작성 항목 및 분류

- (항목) 평균소득, 중위소득
- (인적분류) 성, 연령, 근속기간
- (기업분류) 조직형태, 기업규모, 종사자규모, 산업(중*)분류

* 제조업에 한하여 중분류 수준으로 공표하고, 나머지 산업에서는 대분류 수준

작성 주기 : 기준연도('22) 12월 기준, 연 1회

활용 행정자료

- 4대 사회보험(국민, 건강, 고용, 산재) 및 국세자료(지급명세서·소득신고, 사업자등록, 부가가치세, 법인세 등) 총 30종

작성 방법

- 일자리DB*에 각종 소득 자료를 연계한 후 작성 대상의 모수 추정

* 일자리행정통계 작성을 위해 구축되는 통계용 행정DB

□ 표본설계 배경

- 임금근로자 소득은 현존 행정자료 중 (국세)근로소득지급명세서 및 (국민건강)직장가입자 자료, (고용산재)근로내용확인서에서 파악 가능
- 현행 사회보험 소득은 전수 입수되고 있으나, 국세 소득은 국세청과 별도 협의로 정해진 적정 규모의 표본으로만 입수 가능
- 임금근로자 소득을 정확하게 파악하기 위해서는 사회보험 가입자 뿐만 아니라 미가입자의 국세 소득도 입수하여 반영할 필요

□ 모집단 및 표본추출틀

- (모집단) '22년 12월에 하루 이상 일한 일용근로자 중 사회보험에 미가입하여 소득이 관측되지 않는 자
- (표본추출틀) 일자리DB에서 12월에 일한 일용근로자를 추출한 후 사회보험자료를 연계하여 가입 여부 식별
 - 단위는 일자리(개인×기업체)로, 일자리행정통계는 가구조사와 달리 복수 일자리에서 일한 개인을 각 일자리별로 파악
 - 12월 일용근로자 약 311만 건 중 4대 사회보험 미가입자는 약 146만 명으로 46.8%이며 전년대비 8.3%p 감소

<표1> 종사상지위별 사회보험 미가입 현황

	`20 년(A)			`21 년(B)			증감(B-A)	
	작성대상	미가입	비율(%)	전체	미가입	비율(%)	미가입	비율(%p)
전체	23,311,025	2,175,309	9.3	23,678,605	2,094,906	8.8	-80,403	-3.7
상시	20,045,243	585,580	2.9	20,562,030	637,558	3.1	51,978	8.9
일용	3,265,782	1,589,729	48.7	3,116,575	1,457,348	46.8	-132,381	-8.3

□ 표본규모

- 상대표준오차 3.5% 목표로 산업분류, 종사자규모, 기업규모, 성, 연령대 등 공표 분류별 소득 항목 작성이 가능하도록 산정

1. 모집단 분석

- (표본추출틀) '22년 12월 현재 (개인×기업체) 중 4대보험 미가입 약 146만 건
 - (산업분류) 건설업(37.1%), 보건업 및 사회복지사업(13.8%), 사업시설 관리지원 및 임대(12.5%)가 전체의 약 63.4%를 차지
 - (종사자규모) 50인 미만이 53.0%로 가장 많고, 300인 이상 25.8%, 50~300인 미만 21.3%
 - (성·연령) 성별로는 남녀 약 6:4의 비율을 보이고, 연령대별로는 비교적 고르게 분포하고 있음
 - 여성의 비율이 70대 이상에서 상대적으로 높음
 - (근무개월수) 3개월 이하가 약 46.0%를 차지하고 있고, 여성의 비율이 근무개월 수가 많은 구간(10개월 이상)에서 상대적으로 높음

2. 층화 및 분류지표

□ 층화변수

- 인적 특성인 성, 연령에 비해 산업분류, 종사자 규모의 분포가 편중이 심해 산업분류×종사자규모를 층화변수로 선정
 - 산업분류*(45개)×종사자규모(3개): 총 125개 층**
 - * 제조업은 중분류, 나머지 산업에서는 대분류 수준에서 공표
 - ** 일부 산업분류 및 종사자규모에서 해당 층이 존재하지 않음

□ 분류지표

- 층화 변수 이외에 주요 공표단위인 성, 연령과 근무개월 수 활용
 - * 성: 남, 여
 - * 연령: 10대 이하, 20대, 30대, 40대, 50대, 60대, 70대 이상
 - * 근무개월수: 1~3월, 4~6월, 7~9월, 10~12월

3. 표본규모

□ 상대표준오차 검토

- '21년 기준 결과에 대한 평균소득 상대표준오차(이하 RSE) 시산 결과, 전체 RSE는 약 1.0%, 종사자규모별 RSE는 1.4~2.3%로 나타남

<표2> '21년 기준 전체 및 종사자규모별 평균소득 RSE

(단위: 건, %)

	전체		50인 미만		50~300인 미만		300인 이상	
	표본수	RSE	표본수	RSE	표본수	RSE	표본수	RSE
전체	23,585	1.0	11,007	1.4	6,389	1.9	6,189	2.3

□ 표본규모

- 전년 조사결과와 자료입수 등을 고려하여 '산업분류×종사자규모별' 평균소득 추정을 위해 필요한 표본규모 산출결과 **22,905개** 표본 필요
 - '21년 조사결과 이용, 층별 평균소득 목표 RSE 3~15% 적용
 - * '21년 조사결과 평균소득 RSE 15% 이상(목표RSE 15%), 10~15%(목표RSE 10%), 그 외 전년 수준
 - 각 층별 최소표본수(20개)를 설정하여 적정 표본수 확보

$$n = \frac{n_{old} \left[\frac{RSE(\bar{y})_{old}}{RSE(\bar{y})_{new}} \right]^2}{1 + \frac{n_{old}}{N} \left[\frac{RSE(\bar{y})_{old}}{RSE(\bar{y})_{new}} \right]^2}$$

- $RSE(\bar{y})_{old}$: 직전 조사자료 상대표준오차
- $RSE(\bar{y})_{new}$: 목표 상대표준오차

4. 표본추출

□ 표본추출

- 추출방법: 층화계통추출
 - 층별 분류지표(성, 연령, 근무개월수)로 정렬한 후 계통추출
 - * 성: 남, 여 * 연령: 10대 이하, 20대, 30대, 40대, 50대, 60대, 70대 이상
 - * 근무개월수: 1~3월, 4~6월, 7~9월, 10~12월

□ 표본추출 결과

- 산업분류×종사자규모별 총 22,905개 표본 추출
- 표본추출틀과 표본 분포는 근사적으로 동일한 분포를 나타냄

1. 설계가중값

- 층별 추출률(대표성)이 다르기 때문에 자료 분석 및 통계작성을 위해서는 반드시 가중값 적용이 필요
- 설계층 내 표본단위들의 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(산업분류*종사자규모)
- i : 표본단위
- N_h : h 층의 모집단 수
- n_h : h 층의 표본 수

<표8> 설계가중값 분포

구분	표본수	최소값	Q1	중앙값	평균값	Q3	최대값	표준편차	합계
일용	22,905	1.0	11.9	30.8	63.6	110.0	236.1	67.1	1,457,348

2. 사후 가중값

- 표본추출 시 모집단 정보가 확정되어 사후보정하지 않음
 - '19년 기준 표본설계 시부터 설계가중값만 작성

3. 추정식

○ 총합 및 평균 추정량

총합 추정량	$\hat{Y} = Y_c + \hat{Y}_s = \sum_{i=1}^{N_c} y_i + \sum_{h_s=1}^{H_s} \sum_{j=1}^{n_{hs}} y_{hsj} w_{hsj}$	
평균 추정량	$\hat{\mu} = \hat{Y}/w_{..}$	
첨자	c : 전수(일반근로)	변수 y : 관심변수
	s : 표본(일용근로)	w : 가중값
	h : 설계층(1, ..., H)	n : 표본수
	i, j : 표본단위	N : 모집단수

○ 분산 추정량

총합 추정량의 분산추정량	$\hat{V}(\hat{Y}) = \sum_{h=1}^H \hat{V}_h(\hat{Y}_s)$ $\hat{V}_h(\hat{Y}_s) = \frac{n_{hs}(1-f_{hs})}{n_{hs}-1} \sum_i^{n_{hs}} \left\{ w_{hsi} y_{hsi} - \frac{1}{n_{hs}} \left(\sum_i^{n_{hs}} w_{hsi} y_{hsi} \right) \right\}^2$ $f_{hs} = n_{hs}/N_{hs}$
평균추정량의 분산추정량	$\hat{V}(\hat{\mu}) = \hat{V}(\hat{Y})/(w_{..})^2$ $\hat{V}_h(\hat{\mu}_s) = \hat{V}_h(\hat{Y}_s)/(w_{hs.})^2$

○ 표준오차 및 상대표준오차

총합 추정량	$SE(\hat{Y}) = \sqrt{\hat{V}(\hat{Y})}$ $RSE(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} \times 100$
평균추정량	$SE(\hat{\mu}) = \sqrt{\hat{V}(\hat{\mu})}$ $RSE(\hat{\mu}) = \frac{SE(\hat{\mu})}{\hat{\mu}} \times 100$