

발 간 등 록 번 호

11-1240000-000757-01

2015년도 국가통계 품질개선 컨설팅 연구용역

『소프트웨어기술자임금실태조사』 품질개선 컨설팅 최종결과보고서

2015. 11.

주 의

1. 이 보고서는 통계청에서 수행한 국가통계 품질개선 컨설팅 연구용역 결과보고서입니다.
2. 이 보고서에 대한 지식재산권 귀속 등에 대하여는 「용역 계약 일반조건」 제35조의 2(계약목적물의 지식재산권 귀속 등) 및 「일반용역계약특수조건」 제16조(계약목적물의 지식재산권 귀속 등)에 의합니다.

제 출 문

제 출 문

통계청장 귀하

본 보고서를 「소프트웨어기술자임금실태조사」 품질개선
컨설팅의 최종결과 보고서로 제출합니다.

2015년 11월

한국조사연구학회



연구진

책 임 연 구 원	이기성, 우석대학교 아동복지학과 교수
공 동 연 구 원	손창균, 동국대학교 응용통계학과 조교수
연 구 보 조 원	강혜진, 동국대학교 응용통계학과 석사과정 수료

주요 자문위원

박종모, 정보통신산업진흥원 부설 소프트웨어공학센터 수석연구원

최종결과보고서 요약문

연구과제명	소프트웨어기술자임금실태조사 품질개선 컨설팅
주제어	소프트웨어기술자임금실태조사, 층화, 표본배분, 가중치
연구기간	2015. 9. 1. - 2015. 11. 30.
연구기관	한국조사연구학회
연구진구성	이기성, 우석대학교 아동복지학과 교수 손창균, 동국대학교 응용통계학과 조교수

본 연구에서는 현행 「소프트웨어기술자임금실태조사」 결과와 표본설계 그리고 최신의 모집단 분석을 통해 소프트웨어기술자의 노임단가를 정확히 공표하기 위한 새로운 표본설계 방안(모집단의 정의, 층화, 표본크기 결정 및 표본배분, 표본추출방법, 가중치 부여 및 모수추정, 이상치 처리 방안 등)을 마련하였다.

「소프트웨어기술자임금실태조사」는 1996년 한국소프트웨어산업협회 주관으로 첫 조사를 실시하여 1998년에 통계청으로부터 승인을 받아 매년 실시해 오고 있으며, SW사업을 영위하는 기업체 내 SW기술자의 실지급 임금을 조사하여, 결과를 SW사업 수행시 투입 기술자의 노임단가로 적용하는 중요한 조사이다.

본 연구의 주요 결과를 정리하면 다음과 같다.

첫째, 2015년 표본설계 방법을 검토하였고, 2014년 소프트웨어기술자임금실태조사 원시자료를 이용해서 주요 문항별 특성을 분석하였다. 그리고 주요 조사항목의 추정결과에 대한 상대표준오차를 계산하여 추정의 정확도를 평가하였다. 소프트웨어기술자의 노임단가 등의 주요 변수에 대한 등급별 추정결과의 통계적 정확도는 대체로 안정적인 수준이었으나, 추정의 정확도가 떨어지는 경우가 부분적으로 나타나고 있었다.

둘째, 목표모집단과 조사모집단의 정의를 검토하여, 조사모집단을 구체적으로 정의하였고, 최신 모집단 자료를 확보한 후 모집단의 특성을 분석하여 새로운 표본설계안에 반영하였다.

셋째, 현행 조사결과와 모집단 분석을 통해 표본크기를 결정하였고, 다양한 표본배분방법을 제시하였다.

넷째, 표본추출방법, 가중치 작성 및 모수 추정방법, 이상치 처리 방안 등을 연구하여 새로운 표본설계방안을 제시하였다.

차 례

제 1 장 서론	1
1. 연구배경 및 필요성	1
2. 연구내용 및 방법	2
3. 표본설계의 기본 목표와 절차	4
제 2 장 2014년 조사데이터 분석 및 2015년 표본설계 검토	6
1. 2015년 표본설계 검토	6
2. 2014년 조사데이터 분석	11
3. 2015년 표본설계 및 2014년 조사데이터의 한계점	17
제 3 장 새로운 표본설계	19
1. 새로운 표본설계의 특징	19
2. 모집단 정의 및 분석	19
3. 층화	24
4. 표본크기 결정 및 표본배분	24
5. 표본추출	36
6. 가중치 작성 및 추정	36
7. 이상치 검출 및 처리 방안	50

제 4 장 결론 및 요약	59
참고문헌	63
부록 1. 소프트웨어기술자 등급 분류 기준	65
부록 2. 소프트웨어기술자임금실태조사 조사표	68
부록 3. 개정된 중소기업기본법	73
부록 4. 표본추출 및 모수추정을 위한 SAS 프로그램	75

표 차 례

<표 2-1> 한국SW산업협회 SW사업 유형 분류안	7
<표 2-2> 기업규모별 모집단 특성	7
<표 2-3> 종사자 규모별 매출액별 모집단 특성	8
<표 2-4> 종사자 규모별 사업유형별 모집단 특성	9
<표 2-5> 종사자 규모별 매출액별 표본크기 비율	10
<표 2-6> 종사자 규모별 매출액별 전수조사와 표본조사 현황	10
<표 2-7> 소프트웨어기술자 구분별 평균 임금	11
<표 2-8> 종사자 규모 및 매출액 규모별 소프트웨어기술자 임금평균(1인당)	12
<표 2-9> 2014년 소프트웨어기술자 수준별 임금의 이상치 조정 규모	16
<표 3-1> 기업규모별 모집단 포괄 비율	21
<표 3-2> 기업규모별 모집단 특성	22
<표 3-3> 새로운 종사자 규모별 매출액별 모집단 특성	23
<표 3-4> 종사자 규모별 매출액별 표본크기 현황 : 상대표준오차 5%, 10% 수준 · 26	
<표 3-5> 종사자 규모별 매출액별 표본배분(네이만 배분)	29
<표 3-6> 종사자 규모별 매출액별 표본배분(떡등배분)	30
<표 3-7> 종사자 규모별 매출액별 표본기업체의 소프트웨어기술자 분포	31
<표 3-8> 종사자 규모별 표본기업체 1차 표본배분 현황	33
<표 3-9> 종사자 규모별 표본기업체 2차 표본배분 결과	34
<표 3-10> 종사자 규모별 표본배분 결과 최종안	35
<표 3-11> 추출설계와 가중치	38
<표 3-12> 새로운 종사자 규모별 매출액별 기본 가중치	47

제1장 서론

1. 연구배경 및 필요성

소프트웨어기술자임금실태조사는 SW사업을 영위하는 기업체 내 SW기술자의 실질 임금 조사를 통하여, 결과를 SW사업 수행시 투입 기술자의 노임단가로 적용하기 위하여 실시하고 있다. 이 조사를 통해 SW사업에 종사하는 기술 인력의 임금동향을 파악할 수 있다.

소프트웨어기술자의 노임단가를 파악하기 위해 한국소프트웨어산업협회에서는 구 '소프트웨어개발촉진법' 제10조 및 동법시행령 16조에 근거하여 1996년 첫 조사를 실시하였다. 1998년 5월에 통계청으로부터 통계작성 승인을 받았으며, 2001년 8월에 통계 작성방법을 전수조사에서 표본조사로 변경하게 되었다. 2004년 7월에 주5일 근무제 시행여부 항목이 추가되었고, 2009년 3월에 자료조사원의 노임단가 조사 항목이 추가되었으며, 2013년 5월에 공표일, 지역추가(세종시), SW기술자 등급산정기준이 변경되었다.

소프트웨어기술자임금실태조사는 SW사업자로 신고한 업체 및 한국소프트웨어산업협회 회원사로 등록된 기업체를 조사대상으로 하고 있다. 매년 6월에서 7월 사이에 구조화된 설문지를 이용한 전화, 이메일, 팩스 조사를 통해 5월 중 만근한 SW기술자에게 지급된 월평균 임금(기본급, 제수당, 상여금, 퇴직급여충당금, 법인부담금 등)이 조사되고 있다. 공표범위는 SW기술자 등급별 노임단가이며, 이를 보도자료 배포 또는 한국SW산업협회 홈페이지에 게재하고 있으며 공표일은 매년 8월 31일이다.

소프트웨어기술자임금실태조사는 2006년 국가통계 품질진단을 받았고, 2015년 현재 국가통계 품질진단이 진행되고 있다.

2006년 소프트웨어기술자임금실태조사의 품질진단에서 표본설계와 관련하여 제시된 단기 과제로는 “표본추출과정에서 사용한 층화변수와 무응답시의 대체 과정 등 표본추출에 관한 상세한 정보를 공개할 필요하다”는 것과 “조사모집단에 대한 혼동이 없도록 모집단 정의를 명확히 할 필요가 있다”는 것이었다. 그리고 중장기 개선 과제로는 “추출확률과 무응답 요인 등을 감안하여 동 조사에 적절한 가중치를 개발

할 필요가 있다”는 것이었다.

또한 본 연구진이 2006년 국가통계 품질진단 최종보고서와 2014년 소프트웨어기술자임금실태조사 결과 및 2015년의 표본설계 내용을 살펴본 결과 표본설계에서 모수 추정이 이루어지고 있지 않다는 것이 문제점으로 나타났다.

따라서 이번 소프트웨어기술자임금실태조사의 품질개선 컨설팅 연구에서는 2006년 품질진단에서 나타난 개선사항 뿐만 아니라 모집단의 정의 및 분석, 표본추출방법, 표본배분, 가중치 부여 및 모수추정 방안, 이상치 처리 방안 등을 고려한 새로운 표본설계안을 마련할 필요성이 있다.

2. 연구내용 및 방법

본 연구에서는 소프트웨어기술자임금실태조사의 목적을 충실히 달성하기 위해서 조사의 정확성과 신뢰성을 높이고자 한다. 표본이 모집단을 잘 대표하고, 모집단의 특성이 표본에 반영되도록 설계하며, 과거 표본설계의 한계점이나 문제점 등을 검토한다. 또한, 2014년 소프트웨어기술자임금실태조사 자료 분석을 통해 표본배분을 위한 변수 등을 찾아 새로운 표본설계에 반영한다. 이를 위해서 다음과 같은 단계에 걸쳐서 세부적인 내용을 다룬다.

1) 2014년 소프트웨어기술자임금실태조사 자료 분석 및 2015년 표본설계 검토

2014년에 조사된 소프트웨어기술자임금실태조사 원시자료(raw-data)를 이용해서 주요 문항별 특성을 분석한다. 그리고 2015년의 표본설계 전 과정을 상세히 검토해서 현행 조사의 문제점, 개선내용 등을 종합적으로 파악하여 새로운 표본설계에 반영한다.

2) 모집단 분석

조사 목적에 가장 부합되는 목표모집단과 조사모집단이 정의되었는지를 검토한

후 모집단 명부 확보를 통한 적절한 표본추출틀을 구축한다. 구축된 조사모집단의 특성을 분석하여 표본배분에 활용하면 대표성 있는 표본을 얻을 수 있다. 또한 구축된 모집단 자료는 추정 시 가중값 계산에 사용된다.

3) 새로운 표본설계

새로운 표본설계에서는 2015년의 표본설계에서의 미흡했던 점이나 한계점들을 찾아서 보완토록 한다. 2014년 소프트웨어기술자임금실태조사 자료와 최신 모집단 자료를 심층 분석해서 주요 관련변수를 찾고 이를 표본배분에 반영한다. 최종적으로 연구진, 외부전문가 그리고 한국소프트웨어산업협회 실무진들과의 협의를 통해 표본배분 기준을 정한다.

표본추출은 종사자 규모나 매출액을 층화변수로 사용하고, 표본배분 원칙에 따라 표본추출프로그램(SAS)을 이용하여 배분한다. 각 층별로 기업체를 확률비례추출한 후 표본으로 선정된 기업체에서 소프트웨어기술자를 모두 조사한다.

4) 추정식 유도

새로운 표본설계에서는 2015년의 표본설계와는 달리 모수 추정을 위한 추정치와 추정치의 오차계산을 위해 분산추정식을 유도하고 아울러 가중값 계산과정을 다룬다. 그리고 층별 표본배분을 위한 배분공식을 제시한다.

5) 이상치 처리 방안

이상치 처리 방안을 고찰하여, 본 연구에 적절한 이상치 처리 방안을 제안한다.

상기의 단계적인 세부 내용을 수행하기 위해서 다음과 같은 방법을 사용한다.

(1) 조사연구

최근 통계청에서 생산되고 있는 품질개선 컨설팅 최종결과보고서와 본 조사와 유

사한 기술자 임금실태관련 조사들에 대한 자료를 수집하여 품질개선 컨설팅 내용과 표본설계에 관련된 사항을 분석한다. 그리고 모집단 특성 파악, 표본추출률 확보, 조사항목에 대한 분석, 표본오차의 추정을 위하여 연관된 조사 자료를 수집한다. 또한 추정량, 추정량의 오차를 계산하기 위한 가중값 계산, 표본추출방법, 표본배분 공식, 분산추정식 등에 대한 통계이론들을 살펴본다. 문헌검토 및 자료 수집을 통해서 여러 사례들을 검토한다.

(2) 관련전문가의 의견 수렴 및 검토회의 개최

본 조사를 총괄적으로 담당한 한국소프트웨어산업협회 관련자들과 소프트웨어기술자 임금실태관련 전문가, 통계청 관계자들, 그리고 표본설계 전문가들과 자문회의 등을 통해 수시로 문제점 등을 협의한다.

3. 표본설계의 기본 목표와 절차

새로운 표본설계에서는 다음 사항들을 기본목표로 한다.

- (1) 소프트웨어기술자임금실태를 정확히 파악할 수 있도록 표본설계한다.
- (2) 조사연구의 목적에 가장 부합되는 조사모집단을 설정하고 최신의 표본추출률을 사용하여 표본이 모집단의 특성을 잘 대표할 수 있도록 한다.
- (3) 2015년 표본설계의 장·단점을 분석하고 문제점을 보완해서 조사결과의 정확성과 신뢰성을 높인다.
- (4) 확률추출을 근간으로 해서 특성별 추정이 가능하도록 추정량과 추정량의 분산추정식을 유도한다.
- (5) 조사 후 추정 시 문제가 되는 가중치 작성 방법, 이상치 처리 방법 등을 고려하여 조사결과의 정확성을 높인다.

한편, 효율적인 표본설계를 위해서 다음과 같은 단계적 절차를 밟는다.

- (1) 2015년 표본설계의 장·단점을 분석한다.

- (2) 2014년 소프트웨어기술자임금실태조사 자료로부터 사업유형별, 매출액, 종사자 규모별 특성을 분석하고 표본배분을 위한 변수 등을 찾는다.
- (3) 최신 모집단을 확보하여 상세한 분석을 하여, 본 조사의 목적에 부합되는 특성들을 찾아 표본설계에 반영한다.
- (4) 주어진 조사여건 하에서 층화방법, 추출방법, 표본배분방법 등을 결정한다.
- (5) 가중치와 추정량, 그리고 추정량에 대한 오차공식 등을 유도한다.
- (6) 조사 후 추정 시 문제가 되는 가중치 작성방법에 대하여 살펴본다.
- (7) 조사 결과 이상치가 발생할 경우 그 처리 방법들에 대하여 검토하여 적절한 방안을 마련한다.

제2장 2014년 조사데이터 분석 및 2015년 표본설계 검토

1. 2015년 표본설계 검토

본 절에서는 SW사업을 영위하는 기업체 내 SW기술자의 실지급 임금을 조사하기 위해 작성된 2015년 표본설계 내용을 살펴본다.

1) 모집단 분석

2015년 표본설계에서는 소프트웨어산업진흥법 제24조에 의거하여 한국SW산업협회에 신고한 SW사업자 및 협회 회원사 15,293개를 목표모집단으로 사용하였다. 이때, SW사업자란 SW사업을 하는 자(관련법 제2조 4항)를 총칭하며, 소프트웨어산업진흥법 제24조에 의하여 SW사업자로 신고한 업체를 의미한다. 목표모집단 중 모집단 특성(사업 유형, 매출, 인력 규모)이 누락된 업체를 제외한 10,716개 업체를 조사모집단으로 규정하였다.

SW사업 유형 분류는 현재의 통계청 산업 분류 체계(한국표준산업분류, 특수분류, 품목 분류)가 SW산업을 명확하게 구분하지 못하는 한계에 기인하여 협회에서 자체적으로 마련한 시장 분류 체계를 적용하고 있다. 한국SW산업협회 SW사업 유형 분류안을 살펴보면 다음 <표 2-1>과 같다.

한국SW산업협회 SW사업 유형은 컨설팅, 시스템통합, 아웃소싱, IT솔루션, 정보통신/네트워크, 하드웨어관련, 기타 등 12개로 분류하고 있다.

<표 2-1> 한국SW산업협회 SW사업 유형 분류안

SW사업유형(시장분류)		한국표준산업분류 중 SW부문
1	컨설팅(BPR, ISP EA/ITA 등)	컴퓨터시스템 통합 자문 및 구축 서비스업(62021)
2	시스템통합(SI)	
3	아웃소싱(ITO전략/진단, BPO, 통합유지보수서비스/운영 등)	기타 정보기술 및 컴퓨터운영 관련 서비스업(62090)
4	IT 솔루션(커스터마이징을 포함한 자사SW의 개발, 공급, 판매)	58221~2 시스템/응용 소프트 웨어 개발 및 공급업
5	IT 솔루션유통(패키지/솔루션 단순 유통, 판매)	
6	IT 인프라솔루션(ASP, Saas, 클라우드 컴퓨팅 등 SW임대서비 스 공급)	분류 체계 불분명
7	임베디드(임베디드 소프트웨어 관련)	
8	인터넷서비스(포털, 경매, 쇼핑몰, e-Marketplace 등)	63112 호스팅 및 관련 서비 스업
9	인력지원공급(하도급사업 등)	분류 체계 불분명
10	정보통신/네트워크(설비/네트워크구축 등)	컴퓨터시설 관리업(62022)
11	하드웨어관련(하드웨어 제조/유통 공급)	
12	기타	분류 체계 불분명

기업 규모별, 매출액과 종사자 규모별, 사업유형과 조사자 규모별에 따라 조사모
집단의 특성을 분석하고 있다.

기업규모별 모집단 특성을 살펴보면 다음 <표 2-2>와 같다.

<표 2-2> 기업규모별 모집단 특성

구분	기업체		종사자	
	업체수	%	종사자수	%
대기업	234	2.2	262,700	51.4
중소기업	10,482	97.8	248,027	48.6
전체	10,716	100.0	510,727	100.0

대·중소기업 분류는 중소기업기본법 제2조 및 동법시행령 제3조에 근거하여 분

류하였다. 그 결과 전체 10,716개의 기업체 중 중소기업이 차지하고 있는 비율이 97.8%이었고, 대기업이 차지하는 비율은 2.2%이었다. 반면에 종사자 수를 살펴보면 510,727명 중 대기업이 차지하고 있는 비율이 51.4%로 중소기업의 48.6%보다 많이 나타났다.

현행의 종사자 규모별 매출액별 모집단 특성을 살펴보면 다음 <표 2-3>과 같다.

<표 2-3> 종사자 규모별 매출액별 모집단 특성

구분	10억 미만	10억~ 50억 미만	50억~ 100억 미만	100억~ 500억 미만	500억~ 1,000억 미만	1,000억 이상	계
30인 미만 (%)	4,158 (38.8)	3,569 (33.3)	466 (4.3)	143 (1.3)	0 (0.0)	0 (0.0)	8,336 (77.8)
30~100인 미만 (%)	20 (0.2)	742 (6.9)	544 (5.1)	473 (4.4)	21 (0.2)	0 (0.0)	1,800 (16.8)
100~300인 미만 (%)	0 (0.0)	0 (0.0)	54 (0.5)	265 (2.5)	61 (0.6)	40 (0.4)	420 (3.9)
300인 이상 (%)	0 (0.0)	0 (0.0)	0 (0.0)	29 (0.3)	28 (0.3)	103 (1.0)	160 (1.5)
계 (%)	4,178 (39.0)	4,311 (40.2)	1,064 (9.9)	910 (8.5)	110 (1.0)	143 (1.3)	10,716 (100)

종사자 규모에서는 30인 미만이 차지하는 비율이 77.8%로 높게 나타났으며, 그 다음으로는 30~100인 미만이 16.8%를 차지하는 것으로 나타났다. 매출액에서는 10억~50억 미만이 40.2%로 가장 높게 나타났으며, 그 다음으로는 10억 미만이 39.0%로 높게 나타났다. 종사자 규모별 매출액별에서는 종사자 수가 30인 미만이면서 매출액이 10억 미만인 기업체의 비율이 38.8%로 가장 높게 나타났으며, 그 다음으로는 종사자 수가 30인 미만이면서 매출액이 10억~50억 미만인 기업체의 비율이 33.3%로 높게 나타났다.

보유하고 있는 종사자 수가 작은 업체는 발생하는 매출액의 규모도 작은 경우가 대부분이며, 보유하고 있는 종사자 수가 많을수록 발생하는 매출액의 규모도 큰 것으로 나타났다.

현행의 종사자 규모별 사업유형별 모집단 특성을 살펴보면 다음 <표 2-4>와 같다.

<표 2-4> 종사자 규모별 사업유형별 모집단 특성

(단위: %)

구분	컨설팅	시스템 통합	아웃소싱	IT솔루션	정보통신 / 네트워크	하드웨어 관련	기타	계
30인 미만	7.6	32.7	5.7	9.2	3.2	2.5	1.2	62.2
30~100인 미만	4.4	15.1	1.5	3.7	0.9	1.0	0.4	27.0
100~300인 미만	1.9	4.2	0.2	1.0	0.2	0.2	0.1	7.8
300인 이상	0.7	1.3	0.1	0.4	0.1	0.4	0.1	3.0
계	14.6	53.2	7.5	14.3	4.4	4.2	1.8	100.0

사업유형분류 중 1.0% 미만의 분포를 보인 사업유형은 기타분류로 통합하여 분석하였다. 사업유형에서는 시스템통합이 53.2%로 가장 높게 나타났으며, 그 다음으로 컨설팅이 14.6%, IT솔루션이 14.3%로 높게 나타났다. 종사자 규모별 사업유형별에서는 종사자 수가 30인 미만이면서 시스템통합 기업체의 비율이 32.7%로 가장 높게 나타났다.

2) 층화

2015년 표본설계에서는 근로자 임금에 영향을 주는 다양한 요인 중 종사자 수와 매출액이 주요변수에 해당되어, 관련 변수를 기준으로 층화기준으로 설정하였다.

종사자 수 규모는 4개(30인 미만, 30~100인 미만, 100~300인 미만, 300인 이상)의 층으로 매출액 규모는 6개(10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~500억 미만, 500억~1,000억 미만, 1,000억 이상)의 층으로 구성되었다.

3) 표본크기 결정 및 배분

2015년 표본설계에서는 정밀도를 높이기 위해 보수적으로 1,000개 이상의 표본을 추출하기로 하였다. 종사자 규모 및 매출액 구간 등 사업 규모에 따른 층별 기술자 인력 수에 따라 기업체를 정렬한 후 임의추출하였다. 각 층에 따른 모집단 규모를 고려하여 모집단의 크기가 90개 미만은 전수조사 대상으로 하였고, 90개 이상은 표본조사를 하는 것으로 정하였다. 전수조사 대상자를 제외한 나머지 10,463개 기업체

에 대해 각 층에 네이만배분법을 이용하여 배분하였다. 최종 표본 크기는 1,571개이며, 종사자 규모별 매출액별 표본크기 비율 결과는 다음 <표 2-5>와 같다.

<표 2-5> 종사자 규모별 매출액별 표본크기 비율

(단위 : %)

구분	10억 미만	10억~50억 미만	50억~100억 미만	100억~500억 미만	500억~1,000억 미만	1,000억 이상	계
30인 미만	12.1	31.4	5.9	1.2	0.0	0.0	50.5
30~100인 미만	0.4	14.5	12.8	8.4	0.0	0.0	36.1
100~300인 미만	0.0	0.0	1.3	6.6	1.5	0.6	9.9
300인 이상	0.0	0.0	0.0	0.3	0.5	2.6	3.4
계	12.5	45.9	20.0	16.5	2.0	3.2	100.0

종사자 규모에서는 30인 미만이 차지하는 비율이 50.5%로 높게 나타났으며, 그 다음으로는 30~100인 미만이 36.1%를 차지하는 것으로 나타났다. 매출액에서는 10억~50억 미만이 45.9%로 가장 높게 나타났으며, 그 다음으로는 50억~100억 미만이 20.0%로 높게 나타났다. 종사자 규모별 매출액별에서는 종사자 수가 30인 미만이면서 매출액이 10억~50억 미만 기업의 비율이 31.4%로 가장 높게 나타났으며, 그 다음으로는 종사자 수가 30~100인 미만이면서 매출액이 10억~50억 미만인 기업의 비율이 14.5%로 높게 나타났다.

한편, 종사자 규모별 매출액별 전수조사와 표본조사 현황은 <표 2-6>과 같다.

<표 2-6> 종사자 규모별 매출액별 전수조사와 표본조사 현황

구분	10억 미만	10억~50억 미만	50억~100억 미만	100억~500억 미만	500억~1,000억 미만	1,000억 이상
30인 미만	표본조사	표본조사	표본조사	표본조사	없음	없음
30~100인 미만	전수조사	표본조사	표본조사	표본조사	전수조사	없음
100~300인 미만	없음	없음	전수조사	표본조사	전수조사	전수조사
300인 이상	없음	없음	없음	전수조사	전수조사	표본조사

2. 2014년 조사데이터 분석

2014년 조사결과를 바탕으로 현행 조사의 문제점을 파악하고, 2016년 표본설계에 이러한 내용을 반영한 새로운 표본설계 방안을 제시하기 위해 2014년 조사 자료를 분석하였다. 층 구분은 <표 2-6>과 같이 종사자 규모와 매출액 규모를 이용하여 24개 층을 형성하도록 하였다. 이는 2015년의 표본설계를 반영한 층화 기준으로 소프트웨어기술자 평균 임금을 추정하기 위해 같은 층화 기준을 적용하였다.

1) 소프트웨어기술자 구분별 평균 임금

소프트웨어기술자 구분별 평균임금을 추정한 결과 기술사의 평균임금은 8,588,900원이며, 특급 기술자는 7,917,573원, 고급 기술자는 5,734,378원, 중급 기술자는 4,672,278원, 초급 기술자는 3,956,573원 등으로 추정되었다. 한편 기술 등급별 평균 임금에 대한 상대표준오차(CV)를 살펴보면 임금 수준의 변동이 가장 큰 기술등급은 초급 기술자로 10.7%이었으며, 가장 변동이 적은 기술직은 고급기술자로 4.3%로 나타났다.

<표 2-7> 소프트웨어기술자 구분별 평균 임금

구분	기업체수	인원수	평균임금	표준오차	최소	최대	CV(%)
기술사	35	381	8,588,900	766,090	866,090	18,773,230	8.9
특급 기술자	634	15,511	7,917,573	666,213	1,472,917	14,743,905	8.4
고급 기술자	813	10,737	5,734,378	245,219	1,571,897	12,612,518	4.3
중급 기술자	832	10,923	4,672,278	226,786	1,212,701	7,768,591	4.9
초급 기술자	857	16,740	3,956,573	424,140	971,572	8,199,318	10.7
고급 기능사	31	103	3,581,152	165,297	1,711,600	5,513,263	4.6
중급 기능사	41	358	2,901,427	154,739	1,606,833	4,273,916	5.3
초급 기능사	49	280	2,451,870	227,652	923,434	3,568,620	9.3
DB 입력원	22	197	2,337,194	196,165	918,120	3,952,853	8.4

다음으로 층별 소프트웨어기술자 등급별 평균 임금을 추정한 결과가 <표 2-8>과 같다.

<표 2-8> 종사자 규모 및 매출액 규모별 소프트웨어기술자 임금평균(1인당)

종사자	매출액	기술사		특급 기술자		고급 기술자	
		평균	표준오차	평균	표준오차	평균	표준오차
30인 미만	10억 미만	-	-	5,598,075	168,653	4,542,795	153,696
	10억~50억 미만	7,206,679	1,855,051	5,747,136	117,513	4,596,023	82,208
	50억~100억 미만	-	-	6,019,557	322,471	5,638,764	206,579
	100억~500억 미만	-	-	6,754,282	720,290	5,481,106	532,081
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	-	-	-	-	-	-
30~100인 미만	10억 미만	-	-	6,238,622	52,763	4,367,804	466,936
	10억~50억 미만	6,278,212	495,739	5,598,663	178,096	4,784,865	110,731
	50억~100억 미만	6,749,711	1,380,701	6,182,293	173,123	5,010,797	118,604
	100억~500억 미만	-	-	6,661,864	207,957	5,461,817	166,890
	500억~1,000억 미만	-	-	-	-	6,028,625	0
	1,000억 이상	-	-	-	-	-	-
100~300인 미만	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	-	-	4,610,550	294,928	3,845,928	249,778
	100억~500억 미만	6,469,457	206,315	6,348,050	253,457	5,119,369	209,178
	500억~1,000억 미만	6,381,675	861,360	6,722,723	407,719	5,604,374	436,940
	1,000억 이상	-	-	7,399,859	270,338	5,259,741	543,800
300인 이상	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	-	-	-	-	-	-
	100억~500억 미만	-	-	4,736,815	816,977	3,672,045	367,021
	500억~1,000억 미만	5,757,570	0	6,802,569	278,905	5,512,828	449,427
	1,000억 이상	9,325,913	722,938	8,520,689	717,308	6,360,238	346,262

<표 2-8> 종사자 규모 및 매출액 규모별 소프트웨어기술자 임금평균(1인당)-계속

종사자	매출액	중급 기술자		초급 기술자		고급 기능사	
		평균	표준오차	평균	표준오차	평균	표준오차
30인 미만	10억 미만	3,687,419	111,325	2,543,343	98,789	2,356,667	0
	10억~50억 미만	3,735,310	75,566	2,697,057	47,990	3,538,820	129,766
	50억~100억 미만	4,662,489	198,334	2,573,119	90,579	4,588,428	293,738
	100억~500억 미만	4,105,234	311,806	3,569,092	706,266	-	-
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	-	-	-	-	-	-
30~100인 미만	10억 미만	3,593,151	410,681	3,120,052	235,365	-	-
	10억~50억 미만	3,876,494	95,729	2,817,225	58,101	3,039,311	112,580
	50억~100억 미만	4,137,575	130,158	3,002,143	75,056	4,040,006	334,656
	100억~500억 미만	4,192,487	125,495	3,143,601	116,270	4,036,111	223,093
	500억~1,000억 미만	3,711,706	0	2,447,973	0	-	-
	1,000억 이상	-	-	-	-	-	-
100~300인 미만	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	3,489,965	201,051	3,071,473	240,594	-	-
	100억~500억 미만	4,112,940	172,721	3,017,641	108,172	2,898,807	669,767
	500억~1,000억 미만	4,481,032	317,768	3,606,404	374,614	-	-
	1,000억 이상	3,821,509	363,968	3,636,688	390,591	-	-
300인 이상	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	-	-	-	-	-	-
	100억~500억 미만	3,263,844	175,380	2,292,906	65,151	-	-
	500억~1,000억 미만	4,448,217	242,942	3,898,356	233,755	-	-
	1,000억 이상	5,418,365	340,274	5,244,423	405,101	3,804,359	186,207

<표 2-8> 종사자 규모 및 매출액 규모별 소프트웨어기술자 임금평균(1인당)-계속

종사자	매출액	중급 기능사		초급 기능사		DB 입력원	
		평균	표준오차	평균	표준오차	평균	표준오차
30인 미만	10억 미만	-	-	2,099,767	0	1,461,440	252,446
	10억~50억 미만	2,643,980	81,011	2,334,804	171,358	1,680,126	231,470
	50억~100억 미만	4,029,438	296,497	2,222,869	94,655	-	-
	100억~500억 미만	-	-	2,812,083	0	-	-
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	-	-	-	-	-	-
30~100인 미만	10억 미만	-	-	-	-	-	-
	10억~50억 미만	2,437,571	321,196	2,099,197	109,837	2,659,378	539,638
	50억~100억 미만	3,171,108	213,677	2,192,378	155,891	2,772,353	240,829
	100억~500억 미만	2,975,760	258,799	2,156,238	155,524	2,090,228	219,138
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	-	-	-	-	-	-
100~300인 미만	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	2,451,690	0	1,818,096	0	-	-
	100억~500억 미만	3,412,315	278,229	2,632,874	101,148	2,517,246	333,541
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	-	-	-	-	-	-
300인 이상	10억 미만	-	-	-	-	-	-
	10억~50억 미만	-	-	-	-	-	-
	50억~100억 미만	-	-	-	-	-	-
	100억~500억 미만	-	-	-	-	-	-
	500억~1,000억 미만	-	-	-	-	-	-
	1,000억 이상	3,211,770	0	3,461,239	101,379	2,550,080	0

표본 층에 따른 기술 등급별 소프트웨어기술자의 평균 임금은 층별로 차이가 나타나고 있으며, 기술사를 예를 들어보면 30인 미만의 기업체에 근무하는 기술사의 평균 임금은 7,206,679원인 반면, 30~100인 미만인 경우 매출액이 10억~50억 미만인 기업체에 근무하는 기술사는 6,278,212원으로 차이가 나타났다. 300인 이상 매출액 규모가 1000억 이상인 기업체에 근무하는 기술사는 9,325,913원으로 가장 높았으며, 300인 이상 500억~1,000억 미만인 기업체의 기술사는 5,757,570원으로 가장 낮았다. 마찬가지로 소프트웨어기술자의 평균 임금은 이들이 근무하는 기업체의 규모에 따라 차이가 나타남을 알 수 있다.

2014년 조사 결과를 기업체 매출액 규모별로 보면 중소기업과 대기업의 구분이 모호하며, 실제로 소프트웨어기술자의 임금이 기업체 규모에 따라 차이가 있음을 알 수 있다. 즉, 전체적인 기술 수준에 따른 임금 수준이 기업체 규모에 따라 변동하고 있기 때문에 중소기업 근무자와 대기업 근무자의 임금 수준을 비교하기 위해 서라도 기업체 규모를 보다 명확히 할 필요가 있음을 알 수 있다.

2) 이상치 조정 규모

한국소프트웨어산업협회에서는 조사 과정에서 응답된 임금이 전년 응답값과 비교하여 $\pm 10\%$ 이상 증가 혹은 감소여부를 확인한 후 타당성 여부를 확인하여 이상치로 판단되면 재조사하여 수정하거나 사유를 기재하는 방법으로 이상치를 관리하였다. 이러한 이상치 사후 조정 작업을 거친 데이터는 총 100건으로 기술자 등급별로 그 현황을 살펴보면 <표 2-9>와 같다.

재조사를 하여 임금의 이상치가 가장 많이 조정된 소프트웨어기술자 수준은 초급 기능사로 나타났다. 초급 기능사의 경우 이상치 사후 조정 작업을 거친 데이터는 33건으로 재조사 총 건수 중 33.0%의 비율을 차지하는 것으로 나타났다. 다음으로 중급 기능사가 23.0%, DB 입력원이 11.0%의 순으로 높게 나타났다. 기술사나 기술자에 비하여 기능사에서 임금의 변동 폭이 크게 나타나 이상치로 판단되어 재조사가 이루어지는 비율이 높음을 알 수 있다.

<표 2-9> 2014년 소프트웨어기술자 수준별 임금의 이상치 조정 규모

구 분	빈도	백분율
기술사	3	3.0
특급 기술자	6	6.0
고급 기술자	8	8.0
중급 기술자	7	7.0
초급 기술자	6	6.0
중급 기능사	25	25.0
초급 기능사	33	33.0
DB 입력원	12	12.0
계	100	100.0

2014년의 조사결과를 살펴보면, 이상치 보정과정의 명확한 기준이 없으며, 보정 자료 또한 기존의 자료를 활용하고 있어 적절한 보정이 수행되었는지에 대한 판단이 모호하다. 따라서 향후 객관적인 기준에 따라 이상치 보정이 필요할 것으로 생각된다.

3) 가중치 부여

소프트웨어기술자에 대한 임금 통계산출을 위해서는 표본설계에 따라 기본가중치와 조사과정에서 발생하는 무응답 조정을 위한 무응답 조정 가중치 및 층별 모집단 규모에 따른 사후조정 가중치를 적용한 가중치를 부여함으로써 불균등확률추출에 의한 편향의 축소 및 표본의 대표성 확보를 위해 가중치를 부여한 통계산출이 요구된다.

그러나 2014년 조사결과에서는 가중치를 부여하지 않고 단순 평균을 산출하여 평균 임금을 제시하고 있다. 이는 모집단에 대한 표본의 대표성에 문제가 발생할 수 있으며, 단순 평균은 추정치로서 부적절하기 때문에 추후 이에 대한 개선이 요구된다.

4) 조사표

2014년 조사결과에 대한 원 자료에서는 기업체별 소프트웨어기술자의 수와 그에 대한 평균 임금을 수록하고 있다. 그러나 소프트웨어기술자 1인에 대한 조사 기준 월 임금을 조사함으로써 보다 명확한 통계산출이 가능할 것이다. 즉, 기술자 1인의 조사 기준월 임금 등을 기업체별로 조사함으로써 1단계에서 기업체를 추출하고 2단계에서 표본기업체에 근무하고 있는 소프트웨어기술자들의 임금을 1인당으로 조사함으로써 추정과정에서 이러한 정보를 보다 많이 활용할 수 있기 때문이다.

또한 이와 같이 조사된 결과들은 가중치 산출에도 매우 유용하며, 실제로 기업체 단위로 부여되는 가중치와 소프트웨어기술자 1인에 대한 개인 가중치 산출이 가능하기 때문에 향후 조사표 개선이 필요하다.

3. 2015년 표본설계 및 2014년 조사데이터의 한계점

1) 모집단의 정의

2015년 표본설계에서는 소프트웨어산업진흥법 제24조에 의거하여 한국SW산업협회에 신고한 SW사업자 및 협회 회원사를 목표모집단으로 사용하고 있으며, 목표모집단 중 모집단 특성(사업 유형, 매출, 인력 규모)이 누락된 업체를 제외한 업체를 조사모집단으로 규정하고 있다. 조사명칭이 소프트웨어기술자임금실태조사이므로 목표모집단과 조사모집단에 이 표현이 좀 더 구체적으로 포함될 필요가 있다.

또한 한국SW산업협회에서는 최신의 표본추출틀 구성과 관리를 위하여 지속적인 노력을 기울일 필요가 있다.

2) 종사자 수 규모

2015년 표본설계에서는 4개(30인 미만, 30~100인 미만, 100~300인 미만, 300인 이상)의 종사자 수 규모로 층화를 하고 있다. 모집단 분석 결과 30인 미만이 차지하는 비율이 77.8%로 다른 종사자 규모 층에 비하여 너무나 높은 비율을 차지하는 것으로 나타났다. 따라서 30인 미만의 종사자 수 규모 층을 좀 더 세분화 할 필요가 있다.

3) 매출액 규모

2015년 표본설계에서는 6개(10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~500억 미만, 500억~1,000억 미만, 1,000억 이상)의 매출액 규모로 층화를 하고 있다. 500억~1,000억 미만, 1,000억 이상으로 층화를 할 경우 대기업과 중소기업의 매출액 기준으로 구분한지 못하는 문제점이 있다. 현재 소프트웨어 산업은 매출액 800억원을 기준으로 대기업과 중소기업을 분류하고 있으므로 매출액 규모에 800억원이 포함되어 층화할 필요가 있다.

4) 표본크기 산출식

2015년 표본설계에서는 표본배분과 관련된 식만을 제공하고 있을 뿐 표본크기를 산출하는 식을 제공하고 있지 않다. 표본크기 결정 식이 제시될 필요가 있다.

6) 가중치 산정 및 모수 추정

한국소프트웨어산업협회에서는 지금까지 가중치를 부여하지 않고 단순 평균을 산출하여 평균 임금을 제시하고 있다. 표본의 대표성 확보를 위하여 가중치의 산정이 필요하며, 이를 이용하여 모수추정이 이루어져야 한다.

7) 이상치 판단 기준

한국소프트웨어산업협회에서는 조사 과정에서 응답된 임금이 전년 응답값과 비교하여 $\pm 10\%$ 이상 증가 혹은 감소여부를 확인한 후 타당성 여부를 확인하여 이상치로 판단되면 재조사하여 수정하거나 사유를 기재하는 방법으로 이상치를 관리하고 있다. 이상치가 $\pm 10\%$ 이상 증가 혹은 감소라는 주관적인 견해에 의해 판단되고 있어, 이상치 보정과정의 명확한 기준이 없는 것으로 나타났다. 따라서 이상치에 대한 객관적인 기준 마련이 필요하며, 이에 따라 이상치를 보정하는 방안 등이 필요하다.

제3장 새로운 표본설계

1. 새로운 표본설계의 특징

소프트웨어기술자임금실태조사를 위한 표본설계가 갖는 특징은 다음과 같다.

- (1) 표본규모는 전년도 수준을 유지하되, 조사의 목적을 충실히 달성하기 위해서 합리적인 표본배분 원칙과 실사의 효율성을 높이도록 한다.
- (2) 새로운 표본설계에서는 목표모집단과 조사모집단의 정의를 검토하여, 조사모집단을 구체적으로 정의하고자 한다.
- (3) 최신의 모집단 분석 결과와 2014년 조사데이터 분석결과를 토대로 표본배분을 하는데 네이만배분(Neyman allocation)과 멱등배분(Power allocation) 등을 검토하고, 조사여건을 고려하여 최종 결정한다.
- (4) 본 설계에서는 가중치 부여 과정을 포함해서 모수 추정식과 추정식의 오차계산 공식을 유도한다.

2. 모집단 정의 및 분석

1) 모집단의 정의

소프트웨어기술자임금실태조사의 목표모집단은 '소프트웨어산업진흥법 제24조에 의거하여 한국SW산업협회에 신고한 SW사업자 및 협회 회원사 15,293개사'로 정의하고 있다. 목표모집단은 2013년 SW사업자로 신고한 기업체(12,471개사)와 본회 회원으로 등록된 기업체(1,218개사) 그리고 2014년 임금실태조사 대상 기업체(1,604개사)로 구성되어 있다. 이 때, SW사업자란 SW사업을 하는 자(관련법 제2조 4항)를 총칭하며, 소프트웨어산업진흥법 제24조에 의하여 SW사업자로 신고한 업체를 말한다. 또한 소프트웨어산업진흥법 제2조(정의)에서 '소프트웨어사업자'란 소프트웨어사업을 하는 자를 말한다.

그리고 조사모집단은 'SW사업자로 신고한 업체 및 본 협회 회원사로 등록된 기

업체 10,716개사'라고 정의하고 있다. 이 때, 조사모집단은 목표모집단 15,293개 업체 중에서 모집단 특성을 나타낼 수 있는 정보(사업 유형, 매출, 인력 규모)가 있는 업체 10,716개 업체로 정하였다.

목표모집단과 조사모집단의 정의를 비교해 보면, 거의 차이가 없이 동일하게 표현되어 구체적인 조사모집단의 정의가 필요한 것으로 나타났다. 이는 2006년 소프트웨어기술자임금실태조사의 품질진단에서도 보완이 필요한 사항으로 지적된바 있다. 그리고 본 조사의 목적인 소프트웨어기술자의 노임단가를 파악하기 위해서는 현행 모집단의 정의에서의 'SW사업자 및 협회 회원사'라는 표현보다는 'SW기업체에 종사하는 소프트웨어기술자'라는 표현으로 수정할 필요가 있다. 또한 소프트웨어기술자임금실태조사 결과를 이용하여 기술자의 노임 단가를 공표할 때, 8등급으로 구분되는 소프트웨어기술자와 DB 입력원의 노임단가를 함께 발표하고 있으므로 모집단을 표현하는데 있어서 'SW기업체에 종사하는 소프트웨어기술자'라는 표현보다는 'SW기업체에 종사하는 소프트웨어기술자 및 DB 입력원'으로 구체적인 내용을 기술할 필요가 있다.

그래서 이러한 내용을 바탕으로 본 조사를 총괄적으로 담당한 한국소프트웨어산업협회 관련자들과 소프트웨어기술자 임금실태관련 전문가들, 통계청 관계자들의 의견을 모아 목표모집단과 조사모집단 및 표본추출틀을 새롭게 정의하였다.

결론적으로 소프트웨어기술자임금실태조사를 위한 목표모집단과 조사모집단 및 표본추출틀은 다음과 같다.

목표모집단	소프트웨어산업진흥법 제24조에 의거하여 한국SW산업협회에 신고한 SW기업체에 종사하는 소프트웨어 기술자 및 DB입력원
조사모집단	한국SW산업협회에 신고한 SW기업체 중 사업유형, 종사자 수 및 매출액 등이 누락된 기업체를 제외한 나머지 18,939개 기업체에 종사하는 소프트웨어 기술자 및 DB입력원
표본추출틀	조사년도 기준 가장 최근 시점에 한국SW산업협회에 신고한 SW기업체 리스트

2015년 현재 기업체 모집단 리스트 22,031개 업체 중에서 매출액 또는 종사자 수

에 대한 정보가 없는 기업체는 316개 업체이었고, 이 업체를 제외한 21,715개 기업체에 대해 주요 업종이 SW기술개발 업종인 기업체는 18,939개 업체로서 목표모집단 대비 조사모집단의 포괄성은 86%이며, 전체적으로 목표모집단 대비 조사모집단의 포괄성을 높지만, 각 층별 포괄성은 기업체의 특성에 따라 다양한 포괄성을 나타내고 있다. 10억 미만의 1~4인 기업체의 포괄성은 66.5%이며, 매출액 10억 미만 중 종사자 50~99인 미만은 50%로 나타났다. 반면 매출액 50억~100억 미만에서 종사자 1~4인 미만은 100%의 포괄성을 나타내고 있어 전체적으로 매출액 규모가 50억 미만이고, 종사자 규모가 50인 이상의 경우 포괄성이 떨어지는 것으로 나타났다.

기업체 매출액 규모가 작은 층의 포괄성이 매출액 규모가 큰 기업체의 포괄성에 비해 상대적으로 낮게 나타나고 있어 소규모 기업체의 포괄성을 높이기 위한 모집단 관리가 필요한 것으로 나타났다. 즉, 소규모 기업체는 기업체의 특성상 휴폐업과 업종 전환과 같은 변동이 상대적으로 규모가 큰 기업체에 비해 크기 때문에 소규모 기업체에 대한 관리가 지속적으로 이루어져야 할 것이다.

<표 3-1> 기업규모별 모집단 포괄 비율

(단위:%)

구분	10억 미만	10억~50억 미만	50억~100억 미만	100억~300억 미만	300억~800억 미만	800억 이상	계
1~4인	65.5	97.1	100.0	100.0	0.0	-	67.6
5~9인	84.8	98.4	98.9	100.0	-	100.0	88.8
10~19인	87.8	97.5	97.9	97.3	100.0	-	93.9
20~49인	83.5	97.7	98.3	97.5	100.0	100.0	96.6
50~99인	50.0	95.8	97.4	97.8	100.00	100.0	96.3
100~299인	11.1	88.2	96.1	97.7	97.6	100.0	96.3
300인 이상	0.0	50.0	100.0	100.0	96.3	98.5	96.7
계	76.1	97.7	98.0	97.7	98.0	99.1	86.0

2) 기업규모별 모집단 특성

기업규모별 모집단 특성을 살펴보면 중견기업은 대기업군으로 포함시키고, 나머지는 대기업과 중소기업으로 구분하면 다음 <표 3-2>와 같다.

<표 3-2> 기업규모별 모집단 특성

구분	기업체		종사자	
	업체수	%	종사자수	%
대기업	470	2.5	595,720	61.3
중소기업	18,469	97.5	376,268	38.7
전체	18,939	100.0	971,988	100.0

대·중소기업 분류는 중소기업기본법 제2조 및 동법시행령 제3조에 근거하여 분류하였다. 그 결과 전체 18,939개의 기업체 중 중소기업이 차지하고 있는 비율이 97.5%이었고, 대기업이 차지하는 비율은 2.5%이었다. 반면에 종사자 수를 살펴보면 971,988명 중 대기업이 차지하고 있는 비율이 61.3%로 중소기업의 38.7%보다 많게 나타났다.

3) 종사자 규모별 매출액별 모집단 특성

2014년의 종사자 규모별 매출액별 모집단 특성을 살펴보면, 종사자 규모에서는 30인 미만이 차지하는 비율이 77.8%로 다른 규모에 비해 너무 크게 나타나고 있어 종사자 규모의 층을 다시 조정할 필요가 있다. 4개의 층으로 구성된 층을 좀 더 세분화하여 종사자 규모의 층을 ICT 통계에서 주로 활용하는 1~4인, 5~9인, 10~19인, 20~49인, 50~99인, 100~299인, 300인 이상으로 분류하는 것이 바람직하다.

또한 매출액 규모에서도 50억 미만이 전체 79.2%를 차지하고 있었고, 50억~100억 미만이 9.9%인 것으로 나타났다. 소프트웨어 산업통계(2015)에서는 매출액 규모를 10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~300억 미만, 300억~5,000억 미만, 5,000억 이상으로 분류하고 있다. 소프트웨어 산업통계(2015)의 매출액 규모의 기준을 고려하여 소프트웨어기술자임금실태조사에서 사용하고 있는 매출액 규모를 재조정할 필요가 있다. 소프트웨어기술자임금실태조사의 매출액 100억~500억 미만과 500억~1,000억 미만의 층들을 소프트웨어 산업통계(2015)의 매출액 기준과 중소기업 매출액 기준 등을 고려하여 100억~300억 미만, 300억~800억 미만, 800억 이상으로 재조정하고자 한다. 2015년부터 새롭게 적용되는 중소기업기본법(부록 3)에 따르면, 근로자 수나 자본금 등의 지표로 중소기업 여부를 판단하던 기존의 복잡한 방식에서 3년 평균 매출액만을 판단기준으로 보는 간단한 방식으로 개정되었다. 따라

서 소프트웨어 산업의 경우 평균매출액 등이 800억 이하인 경우 중소기업으로 분류되게 되었다. 이러한 사실을 토대로 매출액에 규모를 구분하는데 800억을 포함하기로 결정하였다. 매출액 규모를 정리하면 10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~300억 미만, 300억~800억 미만, 800억 이상으로 6개의 층으로 분류해 볼 수 있다.

결론적으로 새로운 표본설계에서는 종사자 규모를 7개 층(1~4인, 5~9인, 10~19인, 20~49인, 50~99인, 100~299인, 300인 이상)으로 매출액 규모를 6개 층(10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~300억 미만, 300억~800억 미만, 800억 이상)으로 분류하는 것으로 결정하였다.

새로운 종사자 규모별 매출액별 기준에 따라 모집단의 특성을 분석해 보면 다음 <표 3-3>과 같다.

<표 3-3> 새로운 종사자 규모별 매출액별 모집단 특성

구분	10억 미만	10억~ 50억 미만	50억~ 100억 미만	100억~ 300억 미만	300억~ 800억 미만	800억 이상	계
1~4인 (%)	3,661 (19.3)	370 (2.0)	12 (0.1)	1 (0.0)	0 (0.0)	0 (0.0)	4,044 (21.4)
5~9인 (%)	3,672 (19.4)	1,685 (8.9)	89 (0.5)	15 (0.1)	0 (0.0)	1 (0.0)	5,462 (28.8)
10~19인 (%)	1,531 (8.1)	2,398 (12.7)	323 (1.7)	73 (0.4)	5 (0.0)	0 (0.0)	4,330 (22.9)
20~49인 (%)	233 (1.2)	1,846 (9.7)	711 (3.8)	353 (1.9)	36 (0.2)	3 (0.0)	3,182 (16.8)
50~99인 (%)	12 (0.1)	229 (1.2)	331 (1.7)	351 (1.9)	95 (0.5)	10 (0.1)	1,028 (5.4)
100~299인 (%)	1 (0.0)	15 (0.1)	73 (0.4)	257 (1.4)	162 (0.9)	96 (0.5)	604 (3.2)
300인 이상 (%)	0 (0.0)	1 (0.0)	7 (0.0)	26 (0.3)	52 (0.3)	203 (1.1)	289 (1.5)
계 (%)	9,110 (48.1)	6,544 (34.6)	1,546 (8.2)	1,076 (5.7)	350 (1.8)	313 (1.7)	18,939 (100.0)

종사자 규모에서는 5~9인 미만이 차지하는 비율이 28.8%로 높게 나타났으며, 그 다음으로는 10~19인 미만이 22.9%를 차지하는 것으로 나타났다. 매출액에서는 10억 미만이 48.1%로 가장 높게 나타났으며, 그 다음으로는 10억~50억 미만이 34.6%로 높게 나타났다. 종사자 규모별 매출액별에서는 종사자 수가 5~9인 미만이면서 매출액이 10억 미만 기업체의 비율이 19.4%로 가장 높게 나타났으며, 그 다음으로는 종사자 수가 1~4인 미만이면서 매출액이 10억 미만인 기업체의 비율이 19.3%로 높게 나타났다.

3. 층화

적절한 층화변수를 설정해서 추정치의 생성은 물론 표본크기를 정하고 표본배분에 반영해야 한다.

모집단 분석을 통해 층화변수로는 종사자 규모별, 매출액별, 사업 유형별 등을 고려해 볼 수 있다. 이들을 층화변수로 사용하기 위해서는 모집단의 정보가 이들을 어느 정도 반영할 수 있는지를 가늠해야 한다. 또한 여러 변수를 사용하여 층화를 하면 층의 수가 많게 되고 표본수가 작을 경우에는 상당수의 층에 표본이 적게 배분되는 문제가 발생한다.

본 연구의 새로운 표본설계에서는 종사자 규모와 매출액 규모를 층화 기준으로 하여 7개 종사자 규모 층과 6개 매출액 규모 층으로 구분하여 총 42개의 층을 형성하였다.

사업 유형별 등과 같은 특성별 추정 시에는 조사 후 사후층화방법을 적용해서 추정치를 보정해 주는 것이 바람직하다.

4. 표본크기 결정 및 표본배분

1) 표본크기 결정

소프트웨어기술자임금실태조사의 주관심 변수는 소프트웨어기술자의 월평균 급여 변수이고, 핵심적인 통계작성 단위는 종사자 규모별 매출액별 구분이다. 따라서 표본설계에서 고려하는 중요 층화변수는 종사자 규모와 매출액별 구분이다.

표본조사를 계획할 때에는 조사에서 얻고자 하는 추정값의 오차의 한계를 미리 정해주는데, 이것을 목표오차(목표정도, target precision)라고 한다. 목표오차는 절대적인 값으로 정해줄 수도 있고, 상대적인 값으로 정해줄 수도 있다. 본 연구에서는 종사자 규모와 매출액에 따라 추정값에 차이가 크게 나타날 수 있다는 점을 고려하여 상대표준오차를 기준하여 표본크기를 결정하고자 한다.

표본조사에서 표본크기는 가용 조사인력 및 예산, 조사 소요시간 등 조사에 필요한 제반여건을 고려하고, 작성되는 통계의 표본오차 수준을 종합적으로 검토하여

결정된다. 표본크기는 주어진 예산과 조사인력 등의 제한조건에 따라 정해질 수 있다. 본 연구에서도 전체 조사비용과 인력의 한계로 전체 표본크기의 대략적인 범위가 정해질 수 있다.

이 조사의 주요 관심 조사항목은 소프트웨어기술자의 월평균 급여이다. 목표 상대표준오차를 달성하기 위해 필요한 표본의 크기는 다음의 식을 통해 계산되며, 목표 상대표준오차는 잠정적으로 5%, 10%로 정하였다.

$$n_{new} = n_{old} \times \left(\frac{CV_{old}}{CV_{new}} \right)^2$$

여기서 CV_{old} 는 현행조사의 상대표준오차(2014년도 소프트웨어기술자임금실태조사의 상대표준오차), CV_{new} 는 새로운 조사의 목표상대표준오차, n_{old} 와 n_{new} 는 각각 현행조사와 새로운 조사의 표본크기를 나타낸다.

각 통계작성 단위에서 최종 표본크기는 앞서 구한 새로운 조사의 표본크기에 대하여 유한모집단 수정을 통해서 결정되었다. 목표 상대표준오차가 5%, 10%인 경우의 표본크기는 다음 <표 3-4>와 같다. 목표 상대표준오차를 5% 수준에 맞추기 위해서는 1,500개 이상, 10% 수준에 맞추기 위해서는 500개 이상의 표본크기가 필요하다.

또 다른 표본크기를 구하는 방법으로 다음 식을 활용할 수 있다.

$$n = \frac{\sum_{h=1}^L w_h s_h^2}{\sum_{h=1}^L w_h s_h^2 / N + \left(\alpha \sum_{h=1}^L w_h \bar{y}_h \right)^2}$$

L : 층의 수

w_h : h 층의 가중치

s_h^2 : h 층의 표본분산

N : 기업체 총수

\bar{y}_h : h 층의 매출액 평균

$\alpha = \frac{\sqrt{\text{var}(\bar{y}_{st})}}{\bar{y}_{st}}$: 목표 상대표준오차

결과적으로 새로운 표본설계에서는 목표 상대표준오차를 5%로 고려하여 표본크

기를 1,500개 이상 수준으로 정하기로 하였다.

<표 3-4> 종사자 규모별 매출액별 표본크기 현황 : 상대표준오차 5%, 10% 수준

종사자 규모	매출액	모집단	표본크기	
			목표오차 5%	목표오차 10%
1~4인	10억 미만	3,661	166	42
	10억~50억 미만	370	43	11
	50억~100억 미만	12	12	12
	100억~300억 미만	1	1	1
	300억~800억 미만	0	0	0
	800억 이상	0	0	0
5~9인	10억 미만	3,672	117	29
	10억~50억 미만	1,685	107	27
	50억~100억 미만	89	5	1
	100억~300억 미만	15	15	15
	300억~800억 미만	0	0	0
	800억 이상	1	1	1
10~19인	10억 미만	1,531	65	16
	10억~50억 미만	2,398	102	25
	50억~100억 미만	323	100	25
	100억~300억 미만	73	73	21
	300억~800억 미만	5	5	5
	800억 이상	0	0	0
20~49인	10억 미만	233	49	12
	10억~50억 미만	1,846	46	11
	50억~100억 미만	711	44	11
	100억~300억 미만	353	93	23
	300억~800억 미만	36	10	10
	800억 이상	3	3	3
50~99인	10억 미만	12	12	12
	10억~50억 미만	229	45	11
	50억~100억 미만	331	41	10
	100억~300억 미만	351	54	14
	300억~800억 미만	95	33	33
	800억 이상	10	8	8
100~299인	10억 미만	1	1	1
	10억~50억 미만	15	15	15
	50억~100억 미만	73	73	50
	100억~300억 미만	257	32	8
	300억~800억 미만	162	43	11
	800억 이상	96	22	6
300인 이상	10억 미만	0	0	0
	10억~50억 미만	1	1	1
	50억~100억 미만	7	7	7
	100억~300억 미만	26	9	2
	300억~800억 미만	52	16	4
	800억 이상	203	25	6
계		18,939	1,494	500

2) 표본배분방법

각 층에 따른 모집단 크기를 고려하여 층별 기업체의 수가 90개 미만인 경우는 조사의 여건을 고려하여 2015년의 표본설계와 동일하게 전수조사 층으로 하였고, 나머지 층은 표본조사 층으로 다양한 배분방법에 의해 표본 기업체 수를 정하였다.

따라서 전수층 612개를 제외한 나머지 층에 표본기업체를 배분하는 방법으로 네이만배분과 멱등배분을 고려하였다. 표본배분 공식은 다음의 식(1)과 식(2)를 사용하였다.

(1) 네이만배분

$$n_h = n \times \frac{N_h S_h}{\sum_h N_h S_h} \quad (1)$$

여기서 S_h 는 h 층의 임금의 모집단 표준편차이며, 추정치 s_h^2 을 이용한다.

네이만 배분은 층의 크기와 층내 변동을 함께 고려한 배분방법으로 층의 크기가 크고 층내 변동이 큰 층에 표본을 많이 배분하도록 하는 방법이다. 따라서 네이만 배분은 기업체조사에서 널리 이용되는 배분방법이기도 하다.

(2) 멱등배분

$$n_h = n \times \frac{(N_h S_h)^\alpha}{\sum_h (N_h S_h)^\alpha} \quad (2)$$

여기서 $0 < \alpha \leq 1$ 이다.

멱등배분은 네이만 배분 식에 멱수를 주어 최적배분의 형식을 취하면서, 층의 크기와 표준편차의 곱에 비례하도록 표본을 배분함으로써 특정 층에 지나치게 많은 표본이 배분되는 것을 방지할 수 있는 방법이다. 멱등배분은 네이만 배분의 장점과 비례배분의 장점을 동시에 가지고 있어 최근에 널리 이용되고 있는 배분방법이다. α 값에 따라 다양한 배분이 가능하기에 최적의 α 값을 찾는 데 어려움을 가질 수 있으나, 일반적으로 강제배분을 최소화하는 α 값, 가능하다면 변동계수의 추정치를 최

소화하는 α 값을 최적값으로 선택한다.

일반적으로 대부분의 산업에서 종사자 규모가 5~19인 기업체가 많기 때문에, 산업별 종사자 규모별 층에 네이만 배분을 하게 되면 대부분의 표본기업체가 규모 1층(5-19인)에 배분되어 종사자 규모 300인 이상의 층에는 기업체가 작게 배분되어 표본의 대표성을 크게 훼손할 수 있다. 이러한 문제점을 해결할 수 있는 배분방법으로 역등배분 방법이 활용되고 있다.

목표 상대표준오차와 조건 여건 등을 고려하여 전체 표본크기를 1,500개 업체로 정하였을 때, 종사자 규모별 매출액별 네이만 배분을 한 결과는 다음 <표 3-5>와 같다. 다음으로 종사자 규모별 매출액별 역등배분을 한 결과는 다음 <표 3-6>과 같다.

네이만 배분 결과, 종사자 규모가 1~4인이면서 매출액 규모가 10억 미만인 층에 가장 많은 268개 업체가 배분되었으며, 그 다음으로는 종사자 규모가 5~9인이면서 매출액 규모가 10억 미만인 층에 198개의 업체가 배분되었다.

역등배분 결과, 종사자 규모가 1~4인이면서 매출액 규모가 10억 미만인 층에 가장 많은 표본이 배분되었다. $\alpha=0.3$ 인 경우 113개 업체, $\alpha=0.4$ 인 경우 129개 업체, $\alpha=0.5$ 인 경우 149개 업체, $\alpha=0.6$ 인 경우 170개 업체, $\alpha=0.7$ 인 경우 193개 업체가 배분되었다. 네이만 배분에 비하여 종사자 규모가 1~4인이면서 매출액 규모가 10억 미만인 층에 너무 많은 업체가 배분되는 문제점이 보완되었다.

<표 3-5> 종사자 규모별 매출액별 표본배분(네이만 배분)

종사자 규모	매출액	모집단	표본크기
1~4인	10억 미만	3,661	268
	10억~50억 미만	370	55
	50억~100억 미만	12	12
	100억~300억 미만	1	1
	300억~800억 미만	0	0
	800억 이상	0	0
5~9인	10억 미만	3,672	198
	10억~50억 미만	1,685	88
	50억~100억 미만	89	89
	100억~300억 미만	15	15
	300억~800억 미만	0	0
	800억 이상	1	1
10~19인	10억 미만	1,531	101
	10억~50억 미만	2,398	95
	50억~100억 미만	323	23
	100억~300억 미만	73	73
	300억~800억 미만	5	5
	800억 이상	0	0
20~49인	10억 미만	233	32
	10억~50억 미만	1,846	47
	50억~100억 미만	711	29
	100억~300억 미만	353	29
	300억~800억 미만	36	36
	800억 이상	3	3
50~99인	10억 미만	12	12
	10억~50억 미만	229	17
	50억~100억 미만	331	19
	100억~300억 미만	351	24
	300억~800억 미만	95	0
	800억 이상	10	10
100~299인	10억 미만	1	1
	10억~50억 미만	15	15
	50억~100억 미만	73	73
	100억~300억 미만	257	14
	300억~800억 미만	162	0
	800억 이상	96	0
300인 이상	10억 미만	0	0
	10억~50억 미만	1	1
	50억~100억 미만	7	7
	100억~300억 미만	26	26
	300억~800억 미만	52	52
	800억 이상	203	29
계		18,939	1,500

<표 3-6> 종사자 규모별 매출액별 표본배분(떡등배분)

종사자 규모	매출액	모집단	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$
1~4인	10억 미만	3,661	113	129	149	170	193
	10억~50억 미만	370	69	68	67	66	63
	50억~100억 미만	12	12	12	12	12	12
	100억~300억 미만	1	1	1	1	1	1
	300억~800억 미만	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0
5~9인	10억 미만	3,672	101	114	128	142	156
	10억~50억 미만	1,685	79	83	85	87	89
	50억~100억 미만	89	89	89	89	89	89
	100억~300억 미만	15	15	15	15	15	15
	300억~800억 미만	0	0	0	0	0	0
	800억 이상	1	1	1	1	1	1
10~19 인	10억 미만	1,531	83	87	91	95	98
	10억~50억 미만	2,398	81	85	89	92	94
	50억~100억 미만	323	53	48	44	39	35
	100억~300억 미만	73	73	73	73	73	73
	300억~800억 미만	5	5	5	5	5	5
	800억 이상	0	0	0	0	0	0
20~49 인	10억 미만	233	58	55	51	47	43
	10억~50억 미만	1,846	66	64	62	60	57
	50억~100억 미만	711	57	53	49	45	41
	100억~300억 미만	353	57	53	49	45	40
	300억~800억 미만	36	36	36	36	36	36
	800억 이상	3	3	3	3	3	3
50~99 인	10억 미만	12	12	12	12	12	12
	10억~50억 미만	229	49	43	38	33	28
	50억~100억 미만	331	50	45	40	35	31
	100억~300억 미만	351	54	49	45	40	36
	300억~800억 미만	95	0	0	0	0	0
	800억 이상	10	10	10	10	10	10
100~29 9인	10억 미만	1	1	1	1	1	1
	10억~50억 미만	15	15	15	15	15	15
	50억~100억 미만	73	73	73	73	73	73
	100억~300억 미만	257	42	39	33	28	24
	300억~800억 미만	162	0	0	0	0	0
	800억 이상	96	0	0	0	0	0
300인 이상	10억 미만	0	0	0	0	0	0
	10억~50억 미만	1	1	1	1	1	1
	50억~100억 미만	7	7	7	7	7	7
	100억~300억 미만	26	26	26	26	26	26
	300억~800억 미만	52	52	52	52	52	52
	800억 이상	203	56	53	48	44	40
계		18,939	1,500	1,500	1,500	1,500	1,500

기업체 매출액 및 종사자 규모별로 소프트웨어기술자의 임금 단가를 이용하여 배분한 결과를 살펴보면, 종사자 규모와 소프트웨어기술자 인원 규모 간의 관계가 명

확하지 않기 때문에 단순하게 임금의 분산을 이용하는 것은 한계가 있는 것으로 판단된다.

다음의 <표 3-7>은 층별로 2014년 표본기업체의 소프트웨어기술자 현황을 나타낸 것이다.

<표 3-7> 종사자 규모별 매출액별 표본기업체의 소프트웨어기술자 분포

종사자 규모	매출액 규모	계	표본기업체								
			기술사	특급 기술자	고급 기술자	중급 기술자	초급 기술자	고급 기능사	중급 기능사	초급 기능사	DB 입력원
1~4인	10억 미만	50	0	9	18	11	12	0	0	0	0
	10억~50억 미만	17	0	1	2	3	7	0	2	2	0
	50억~100억 미만	0	0	0	0	0	0	0	0	0	0
	100억~300억 미만	0	0	0	0	0	0	0	0	0	0
	300억~800억 미만	0	0	0	0	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0	0	0	0
5~9인	10억 미만	214	42	47	42	73	1	0	0	9	0
	10억~50억 미만	194	0	15	34	63	75	0	2	5	0
	50억~100억 미만	2	0	0	1	0	1	0	0	0	0
	100억~300억 미만	0	0	0	0	0	0	0	0	0	0
	300억~800억 미만	0	0	0	0	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0	0	0	0
10~19 인	10억 미만	311	0	57	64	87	90	0	0	1	12
	10억~50억 미만	1,206	1	182	281	289	423	9	5	14	2
	50억~100억 미만	1,507	5	137	446	527	389	1	0	2	0
	100억~300억 미만	21	0	0	5	6	9	0	0	1	0
	300억~800억 미만	0	0	0	0	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0	0	0	0
20~49 인	10억 미만	93	0	10	16	27	40	0	0	0	0
	10억~50억 미만	4,967	12	793	947	1,149	1,898	26	61	68	13
	50억~100억 미만	1,822	0	334	367	447	635	5	18	12	4
	100억~300억 미만	463	0	60	97	123	155	5	3	19	1
	300억~800억 미만	0	0	0	0	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0	0	0	0
50~99 인	10억 미만	15	0	3	1	2	9	0	0	0	0
	10억~50억 미만	1,164	0	211	167	263	470	1	6	5	41
	50억~100억 미만	2,989	12	453	608	588	1,106	13	193	7	9
	100억~300억 미만	2,027	0	287	407	493	778	1	18	8	35
	300억~800억 미만	12	0	0	3	3	6	0	0	0	0
	800억 이상	0	0	0	0	0	0	0	0	0	0
100~ 299인	10억 미만	0	0	0	0	0	0	0	0	0	0
	10억~50억 미만	0	0	0	0	0	0	0	0	0	0
	50억~100억 미만	369	0	21	41	77	161	0	17	52	0
	100억~300억 미만	5,092	67	839	1,011	1,185	1,878	10	29	20	53
	300억~800억 미만	878	4	246	170	157	301	0	0	0	0
	800억 이상	384	0	88	117	81	98	0	0	0	0
300인 이상	10억 미만	0	0	0	0	0	0	0	0	0	0
	10억~50억 미만	0	0	0	0	0	0	0	0	0	0
	50억~100억 미만	0	0	0	0	0	0	0	0	0	0
	100억~300억 미만	495	0	40	60	120	275	0	0	0	0
	300억~800억 미만	1,355	2	200	250	321	582	0	0	0	0
	800억 이상	29,504	278	11,472	5,560	4,839	7,238	31	4	64	18
	계	55,151	423	15,505	10,715	10,934	16,637	102	358	289	188

<표 3-7>로부터 표본기업체의 전체 소프트웨어기술자는 55,151명이며, 이를 층별로 살펴보면, 종사자 규모 1~4인 층의 경우 총 67명이며, 5~9인 층은 410명, 10~19인 층은 3,045명, 20~49인 층은 7,345명, 50~99인 층은 6,207명, 100~299인 층은 6,723명, 300인 이상 층은 31,354명으로 종사자 규모가 클수록 소프트웨어기술자의 고용이 많은 것으로 나타나고 있다.

따라서 층별 소프트웨어기술자의 고용 규모를 반영하여 표본기업체를 배분하는 방법을 고려하는 것이 바람직하다. 즉, 종사자 규모에 대한 1차 층화를 고려하여 소프트웨어기술자의 분포를 반영한 1차 표본배분을 실시한 후, 매출액 규모를 고려한 2차 층화에 대해 1차 배분된 표본기업체를 다시 재배정하는 방법을 고려하였다.

네이만 배분의 경우 표본기업체의 층별 소프트웨어기술자 비율(p_h)을 이용하여 다음과 같이 배분하였다.

$$n_h = n \times \frac{N_h \sqrt{p_h q_h}}{\sum_h N_h \sqrt{p_h q_h}}$$

또한 멱등배분의 경우에는 다음과 같은 식을 이용하여 배분하였다.

$$n_h = n \times \frac{(N_h \sqrt{p_h q_h})^\alpha}{\sum_h (N_h \sqrt{p_h q_h})^\alpha}$$

여기서 $0 < \alpha \leq 1$ 이다.

다음의 <표 3-8>은 2014년 표본기업체의 기술자 분포를 고려하여 종사자 규모별 표본기업체를 1차 배분한 결과이다. 배분방법으로는 비례배분, 네이만 배분, 멱등배분을 고려하였다.

2014년 표본기업체의 기술자 분포를 고려한 네이만 배분에서는 종사자 규모 1~4인 층은 63개 업체, 5~9인 층은 210개 업체, 10~19인 층은 443개 업체, 20~49인 층은 485개 업체, 50~99인 층은 146개 업체, 100~299인 층은 89개 업체, 300인 이상 층은 64개 업체가 배분되었다.

한편 멱등배분의 경우 $\alpha=0.3$ 인 경우 1~4인 층은 158개 업체, 5~9인 층은 227개 업체, 10~19인 층은 284개 업체, 20~49인 층은 292개 업체, 50~99인 층은 204개 업체, 100~299인 층은 175개 업체, 300인 이상 층은 160개 업체가 배분되었다. 이와

같이 소프트웨어기술자의 분포를 고려한 표본배분의 경우 소프트웨어기술자의 임금을 고려한 배분보다 층별로 고른 배분결과를 나타내고 있다.

<표 3-8> 종사자 규모별 표본기업체 1차 표본배분 현황

종사자 규모	모집단 기업체 수	소프트 웨어기 술자 수(표본)	비례 배분	네이 만 배분	먹등배분				
					$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$
1~4인	4,044	67	320	63	158	141	125	111	97
5~9인	5,462	410	433	210	227	229	229	228	225
10~19인	4,330	3,045	343	443	284	308	332	356	379
20~49인	3,182	7,345	252	485	292	320	348	376	404
50~99인	1,028	6,207	81	146	204	198	191	183	174
100~299인	604	6,723	48	89	175	162	149	136	123
300인 이상	289	31,354	23	64	160	142	126	110	98
계	18,939	55,151	1,500	1,500	1,500	1,500	1,500	1,500	1,500

최종적으로 종사자 규모와 매출액 규모에 따라 기업체별 소프트웨어기술자 규모를 고려한 네이만 배분과 먹등배분의 결과는 <표 3-9>와 같다.

이 방법은 앞서 배분한 방법과 달리 전수층을 고려하지 않고, 모집단 기업체 전체를 대상으로 하여 표본기업체를 배분하였으며, 표본기업체의 소프트웨어기술자 분포를 고려하여 가능한 많은 기술자들의 임금이 조사될 수 있도록 하였다.

따라서 임금분포보다는 소프트웨어기술자 규모를 고려하여 표본기업체를 배분하는 것이 보다 바람직하며, 층별 기업체의 규모와 소프트웨어기술자 분포를 적절히 감안한 먹등배분을 사용하는 것이 타당할 것이다. 결과적으로 본 조사에서는 표본기업체 규모가 층별로 고르게 배분된 $\alpha=0.4$ 일 때의 먹등배분을 이용하기로 한다.

<표 3-9> 종사자 규모별 표본기업체 2차 표본배분 결과

종사자 규모	매출액	모집단	네이만 배분	떡등배분				
				$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$
1~4인	10억 미만	3,661	57	105	101	95	89	81
	10억~50억 미만	370	6	53	40	30	22	16
	50억~100억 미만	12	0	0	0	0	0	0
	100억~300억 미만	1	0	0	0	0	0	0
	300억~800억 미만	0	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0
5~9인	10억 미만	3,672	144	115	125	132	137	140
	10억~50억 미만	1,685	66	91	91	89	86	82
	50억~100억 미만	89	0	21	13	8	5	3
	100억~300억 미만	15	0	0	0	0	0	0
	300억~800억 미만	0	0	0	0	0	0	0
	800억 이상	1	0	0	0	0	0	0
10~19인	10억 미만	1,531	114	85	94	101	106	110
	10억~50억 미만	2,398	288	113	136	160	185	211
	50억~100억 미만	323	40	62	62	59	56	53
	100억~300억 미만	73	1	23	16	12	9	5
	300억~800억 미만	5	0	0	0	0	0	0
	800억 이상	0	0	0	0	0	0	0
20~49인	10억 미만	233	10	40	34	29	24	20
	10억~50억 미만	1,846	327	113	139	167	197	228
	50억~100억 미만	711	116	83	92	100	106	111
	100억~300억 미만	353	32	57	55	52	49	45
	300억~800억 미만	36	0	0	0	0	0	0
	800억 이상	3	0	0	0	0	0	0
50~99인	10억 미만	12	0	11	7	4	2	2
	10억~50억 미만	229	31	51	50	48	45	41
	50억~100억 미만	331	57	61	64	65	65	63
	100억~300억 미만	351	57	61	62	64	64	63
	300억~800억 미만	95	1	20	15	10	7	5
	800억 이상	10	0	0	0	0	0	0
100~299 인	10억 미만	1	0	0	0	0	0	0
	10억~50억 미만	15	0	0	0	0	0	0
	50억~100억 미만	73	7	33	28	23	19	15
	100억~300억 미만	257	48	58	59	59	58	56
	300억~800억 미만	162	24	47	44	41	37	34
	800억 이상	96	10	36	31	26	22	18
300인 이상	10억 미만	0	0	0	0	0	0	0
	10억~50억 미만	1	0	0	0	0	0	0
	50억~100억 미만	7	0	0	0	0	0	0
	100억~300억 미만	26	3	26	26	19	14	10
	300억~800억 미만	52	11	49	41	34	28	23
	800억 이상	203	50	82	75	73	68	65
계		18,939	1,500	1,500	1,500	1,500	1,500	1,500

3) 표본배분의 최종안

<표 3.9>에서 각 층별로 모집단 기업체가 있으나, 소프트웨어기술자의 비율이 상대

적으로 낮아 표본이 배분되지 못하는 층도 있었다. 또한 모집단의 층별 기업체 규모가 상대적으로 작고, 이들이 고용하고 있는 소프트웨어기술자의 비율이 낮기 때문에 결과적으로 표본배분이 이루어지지 않은 층도 나타나고 있었다.

<표 3-10> 종사자 규모별 표본 결과 최종안

종사자 규모	매출액	모집단	표본크기
1~4인	10억 미만	3,661	101
	10억~50억 미만	370	40
	50억~100억 미만	12	0
	100억~300억 미만	1	0
	300억~800억 미만	0	0
	800억 이상	0	0
5~9인	10억 미만	3,672	125
	10억~50억 미만	1,685	91
	50억~100억 미만	89	13
	100억~300억 미만	15	0
	300억~800억 미만	0	0
	800억 이상	1	0
10~19인	10억 미만	1,531	94
	10억~50억 미만	2,398	136
	50억~100억 미만	323	62
	100억~300억 미만	73	16
	300억~800억 미만	5	0
	800억 이상	0	0
20~49인	10억 미만	233	34
	10억~50억 미만	1,846	139
	50억~100억 미만	711	92
	100억~300억 미만	353	55
	300억~800억 미만	36	0
	800억 이상	3	0
50~99인	10억 미만	12	7
	10억~50억 미만	229	50
	50억~100억 미만	331	64
	100억~300억 미만	351	62
	300억~800억 미만	95	15
	800억 이상	10	0
100~299인	10억 미만	1	0
	10억~50억 미만	15	0
	50억~100억 미만	73	28
	100억~300억 미만	257	59
	300억~800억 미만	162	44
	800억 이상	96	31
300인 이상	10억 미만	0	0
	10억~50억 미만	1	0
	50억~100억 미만	7	0
	100억~300억 미만	26	26
	300억~800억 미만	52	41
	800억 이상	203	75
계		18,939	1,500

소프트웨어기술자 비율을 고려한 역등배분에서 $\alpha = 0.4$ 일 때 층별 표본배분이 고르게 분포되어 있어서, 새로운 표본설계에서는 $\alpha = 0.4$ 일 때 역등배분 결과를 표본배분의 최종안으로 결정하였다.

5. 표본추출

종사자 규모와 매출액 규모에 따른 층에서 매출액에 따라 기업체를 정렬한 후 계통추출을 한다. 종사자 수 규모는 1~4인, 5~9인, 10~19인, 20~49인, 50~99인, 100~299인, 300인 이상으로 7개의 층으로 구성되어 있다. 그리고 매출액 규모는 10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~300억 미만, 300억~800억 미만, 800억 이상으로 6개의 층으로 구성되어 있다.

최종적으로 표본기업체에 근무하는 소프트웨어기술자들의 임금을 파악하는 것이기 때문에 기업체당 소프트웨어기술자의 총수가 파악되어야 하며, 이들의 임금 단가가 조사되도록 해야 한다. 이러한 측면에서 종사자 규모와 매출액 규모로 층화한 후 각 층의 기업체는 집락으로 고려할 수 있을 것이다.

표본 추출은 다음과 같은 절차에 의해 추출한다(부록 참조).

- 1단계) 모집단의 층화 변수에 따라 오름차순으로 정렬한다.
- 2단계) 층화 계통 표본추출을 위해 모집단의 종업수를 MOS(Measure Of Size)로 선정한다.
- 3단계) 최종적으로 배분된 층별 규모에 따라 확률비례 추출한다.
- 4단계) 표본으로 추출된 기업체에 대해 층별로 일련번호(ID)를 부여하여 최종적으로 추출된 표본을 정리한다.

6. 가중치 작성 및 추정

1) 가중치 조정방법

표본조사의 주된 목적은 모집단으로부터 확률추출된 표본자료를 통해 미지의 모

집단의 특성을 추측(inference)하는데 있다. 이때 표본을 추출하는 방법은 다양한 확률표본추출 방법을 적용할 수 있을 것이다. 이와 같이 표본으로 추출된 추출단위들은 모집단에 있는 표본으로 추출되지 않은 단위들을 잘 대표해야 한다. 즉, 표본자료는 모집단에 대한 대표성을 확보해야 하며, 이때 고려할 수 있는 것이 표본자료에 대한 가중(weighting)이다. 즉, 모집단으로부터 크기 n 의 확률표본을 단순임의 추출할 경우 모집단 총합에 대한 추정량 \hat{Y} 는 $N/n \sum y_i$ 로서 i 번째 단위에 대한 승수가 N/n 이 되며 이를 i 번째 단위에 대한 가중치라 한다.

결과적으로 가중의 주된 목적은 표본자료의 모집단에 대한 대표성 확보라 할 수 있다.

이러한 가중은 첫째, 분산과 비용을 감소시킬 수 있으며, 둘째 영역별로 서로 다른 추출률을 할당하여 작은 영역의 추출률을 높임으로서 전체적인 비용을 감소시킬 수 있다. 셋째, 중복, 비포괄성 등의 불완전 표본추출 문제들을 해결할 수 있다. 넷째, 조사에서 발생하는 무응답 문제를 적절히 해결할 수 있다. 그런데, 이와 같은 가중 절차에서 추출단위에 항상 상수의 가중치를 고려할 수 있는데, 모집단 단위가 표본으로 뽑힐 확률이 0이 아닌 값을 갖는 경우로서 이와 같은 가중치를 “자체가중(self weighting)”이라 한다. 이러한 자체가중은 단순임의추출, 층화추출, 확률비례추출, 다단계 추출 등에서 적용할 수 있다. 자체가중방법은 첫째, 조사의 복잡성을 감소시키며, 둘째 잘못된 가중치의 문제를 해결할 수 있고, 셋째, 동일한 표본이 다양한 목적과 다른 조사에서 사용될 수 있는 유연성과 편리성을 가지며, 넷째, 다양한 표본설계에 로버스트(robust)하며, 다섯째 일반적인 가중절차에 비해 이해하기가 쉽다는 장점을 가진다.

일반적으로 조사단위에 대해 가중의 절차는 (1) 추출가중치 또는 기본가중치 (2) 무응답가중치 (3) 사후층화가중치의 단계로 이루어진다.

(1) 기본가중치

기본가중치(base weight)는 표본추출 설계로부터 직접적으로 얻어지는 값이다. 임의의 모집단으로부터 적절한 크기의 표본을 추출할 때 단위가 표본에 포함될 확률의 역수로 기본가중치를 계산할 수 있다. 이때 포함확률은 기지(known)의 값이다. <표 3-11>은 추출설계에 따른 기본가중치를 산정한 것이다.

<표 3-11> 추출설계와 가중치

추출설계	가중치
단순임의추출(비복원)	$w_i = N/n$
단순계통추출(추출간격이 r 인 경우)	$w_i = r$
층화임의추출(전체 L 개의 층)	$w_{hi} = N_h/n_h$
계통추출(불균등확률)	$w_i = M/(nM_i)$
2단계 추출 $p_i^{(1)}$: i 번째 PSU의 추출확률 $p_{ij}^{(2)}$: i 번째 PSU에서 j 번째 SSU의 조건부추출확률	$w_{ij} = 1/(p_i^{(1)} p_{ij}^{(2)})$

(2) 무응답 조정가중치

무응답 조정가중치(nonresponse weight)의 중요한 역할은 조사로부터 발생한 무응답으로 인한 무응답 편향을 제거하기 위한 것이다. 무응답 편향은 무응답자들이 응답자들과 차이가 있을 때 발생하며, 모집단에 대해 매우 높은 비율로 추정치에 영향을 줄때 발생한다. 무응답 조정가중치를 계산하는 방법으로는 이용 가능한 정보의 근원에 따라 '표본에 기초한 무응답조정 방법'과 '외부정보를 이용한 무응답 조정방법'이 있다.

① 표본에 기초한 무응답 조정방법

이용 가능한 정보가 표본으로 한정되며, 전체 모집단에 대한 정보는 알 수 없고, 무응답 단위들의 기본가중치를 표본응답자들에게 배정하여 응답단위들에 대해 조정된 가중치의 합이 전체 표본단위들에 대한 기본가중치의 합이 된다.

$$F_c = \frac{\sum_{i=1}^{n_1} w_i MOS_i + \sum_{i=1}^{n_2} w_i MOS_i}{\sum_{i=1}^{n_1} w_i MOS_i}$$

여기서 n_1 은 응답자들의 수이고, n_2 는 무응답자들의 수이다. w_i 는 기본가중치이고 MOS_i 는 표본추출설계에 의해 결정되는 적절한 크기척도를 나타낸다.

본 조사에서는 MOS를 기업체에 근무하고 있는 전체 기술자의 수를 이용하기로 한다.

그러면 조정값 F_c 를 표본응답자들의 기본가중치를 곱하여 각 단위의 가중치로 고려하면 무응답 조정 가중값 $w_i^{(a)}$ 를 계산할 수 있다.

$$w_i^{(a)} = \begin{cases} F_c w_i, & \text{그룹1} \\ 0, & \text{그룹2} \\ w_i, & \text{그룹3} \end{cases}$$

이때 그룹1은 응답그룹, 그룹2는 무응답그룹, 그리고 그룹3은 부적절한 표본그룹을 나타낸다.

② 외부정보를 이용한 무응답 조정방법

표본에 기초한 무응답 조정 가중치를 계산한 후 외부자료를 이용하여 사후층화, raking, 또는 calibration등의 방법으로 이 가중치를 조정한다. 외부자료를 이용한 경우 조정인자 F_c 가 다음과 같이 수정된다.

$$F_c^* = \frac{MOS_c}{\sum_{i=1}^{n_1} w_i MOS_i}$$

여기서 MOS_c 는 계급 c 에 대해 외부데이터로부터 얻는 크기척도이다.

이때 유일한 외부정보는 표본추출틀로부터 구할 수 있으며 이를 가중계급을 구성

하는데 이용한다. 가중계급을 구성하기 위해 이용 가능한 방법으로는 경험적인 방법, 분류소프트웨어(CHAlD), 성향점수모형, 다중가법회귀나무(MART) 등이 있다.

(3) 사후층화 가중치

① 사후층화 조정

사후층화(post-stratification) 조정은 표본추출틀의 불완전으로 인한 포괄성의 차이, 표본의 불균형 또는 비대표성, 무응답에 의한 차이 등을 조정하기 위해 광범위하게 이용되는 방법이다. 즉, 표본응답자들의 가중치를 조정함으로써 가중된 표본분포가 기지의 모집단분포와 같아지도록 하는 방법이다. 이러한 사후층화 조정을 실시하는 주된 이유는 첫째, 추정치의 정도를 개선할 수 있으며, 둘째, 추정분산을 줄이기 위해 층화와 비추정을 사용할 수 있고, 셋째, 부차모집단간의 포괄성과 무응답에 따른 추정치의 편향을 감소시키며, 넷째, 모집단의 다양한 그룹에 대해 추정치의 일치성을 보장한다.

사후층화를 위한 초기 셀 구성의 원칙은 다음과 같다.

첫째, 모집단 수는 부차 데이터로부터 이용 가능해야 하며 이들의 특성 값들은 관심변수와 양의 상관관계를 가져야 한다.

둘째, 모집단 수는 정확히 목표모집단 수가 되어야 한다.

셋째, 모집단 수는 가중된 조사추정치보다 더욱 신뢰할 수 있어야 한다.

넷째, 하나의 표본응답자는 하나의 셀로 분류될 수 있어야 한다.

다섯째, 표본에 대한 사후층의 정의와 기지의 모집단은 서로 일치해야 한다.

여섯째, 사후층 셀들은 충분히 커야 한다.

② 래킹비 조정

래킹비 조정(raking ratio adjustment)은 전수조사 자료와 표본조사 자료간의 일치성을 확보하기 위해 1940년에 미국에서 Deming과 Stephan에 의해 처음 제안된 방법으로 2차원 분류표상의 각 셀 값을 반복적으로 조정해 가는 방법이다. 래킹비 조정 절차는 기본 가중치를 하나의 주변분포를 이용하여 조정한 후, 두 번째 주변분포를 재차 이용하여 가중치를 조정한다. 이러한 과정을 특정한 수렴조건을 만족할

때까지 반복적으로 수행한다.

래킹비 조정 절차의 단점은 첫째, 반복적으로 셀 값을 조정해가는 과정에서 수렴성을 보장할 수 없다는 점이다. 즉, 래킹해야 할 차원을 잘못 선택할 경우 래킹 과정이 수렴하지 않을 수 있다. 둘째, MSE 계산이 매우 어렵다는 점이다. 왜냐하면 래킹 추정량의 구조가 복잡하기 때문에 직접적으로 MSE를 계산하기 어렵다는 문제가 있다. 셋째, 무응답과 같은 상황에 바로 적용하기란 쉽지 않다.

③ 보정

보정(calibration)은 추출설계에 따른 기본가중치 또는 추출가중치와 보정된 새로운 가중치 간의 차이인 일종의 거리함수(distance function)를 주어진 조건에서 최소로 하는 가중값을 구하는 방법이다. 이때 제한조건을 나타내는 식을 보정방정식(calibration equation)이라 하며 다음과 같이 정의한다.

$$\sum_s w_i x_i = \sum_U x_i = \tau_x$$

여기서 x_i 는 관심변수 y_i 와 강한 상관이 있는 보조변수이다.

보정방정식의 조건하에서 특정한 거리함수를 최소로 하는 새로운 가중치 w_i 를 구하면 된다.

즉, 선형 거리함수 $\sum G(w, d) = \sum (w_k - d_i)^2 / 2d_i$ 를 최소로 하는 원래의 추출 가중치 d_i 의 새로운 가중치 w_i 를 구하면 $w_i = d_i(1 + \sum d_i x_i)(\sum x_i x_i')^{-1}(\tau_x - \hat{\tau}_x)$ 이며 이를 총합추정량의 식에 대입하면 다음과 같은 일반화 회귀추정량(generalized regression estimator : GREG)이 된다.

$$\hat{\tau}_{yGREG} = \hat{\tau}_{yHT} + \hat{B}'(\tau_x - \hat{\tau}_{xHT})$$

이와 같은 회귀추정량의 장점은 연속형 보조정보를 사용할 수 있고 유연성을 가지며, 분산을 정의할 수 있다는 점이다.

(4) 가중의 효과

표본의 추출률과 조사에서 발생하는 무응답, 그리고 프레임의 불완전성으로 인한 비포괄성 문제 등을 해결하기 위해서는 반드시 가중치를 고려해야 한다. 만일 가중을 하지 않으면, 추정치의 편향이 증가하게 되어 추정치의 왜곡을 피할 수 없게 된다. 그러나 가중치를 고려하게 되면 추정치의 편향은 감소하지만 가중의 효과로 인해 분산이 증가하게 된다. 이러한 이유 중의 하나는 각 단위에 부과되는 가중치의 변동이 매우 클 경우에 발생한다.

모집단 평균을 추정함에 있어 분산의 증가분에 기여한 가중치의 효과는 다음과 같은 인자에 의해 측정된다.

$$L = n \times \frac{\sum_h n_h w_h^2}{(\sum_h n_h w_h)^2}$$

여기서 $n = \sum_h n_h$ 로서 실제 표본크기를 나타내며, w_h 는 최종가중치, n_h 는 h 층의 실제 표본크기이다.

이식은 다음과 같이 가중치의 변동계수(CV)로 표현이 가능하다.

$$L = n \times \frac{\sum_j w_j^2}{(\sum_j w_j)^2} = 1 + CV^2(w_j)$$

이때 $CV^2(w_j) = \frac{n}{(\sum_j w_j)^2} \left(\sum_j w_j^2 - \frac{1}{n} (\sum_j w_j)^2 \right)$ 이다.

(5) 가중치의 절단

일단 가중치가 계산되고 불완전성을 보상하기 위해 조정된 다음, 조정된 가중치의 분포를 파악하는 것이 바람직하다. 극히 작은 표본에 의해 극단적으로 큰 가중치는 추정치의 분산을 증가시키는 요인이 된다. 그러므로 가중치의 변동을 고려하여 최대한 수준에서 극단적인 가중치를 절단하는 것이 필요하다. 가중치의 절단

(trimming of weights)은 대체로 무응답에 대한 조정 후에 수행하는 것이 일반적이다.

가중치의 절단 작업은 분산을 축소시키는 효과가 있지만, 다른 한편으로는 추정치의 편향을 야기하게 된다. 따라서 매우 큰 가중치에 대해 절단 작업을 수행하면, 분산의 크기는 줄일 수 있지만, 편향의 크기가 상대적으로 증가하게 되는 문제가 있다. 따라서 가중치의 절단은 절단을 함으로서 발생하는 분산의 감소분보다 총 MSE의 영향이 적도록 가중치를 절단하는 것이 바람직하다.

층화추출설계에 대해, 가중치 절단과정은 각 층 내에서 수행되는 것이 이상적이다. 먼저 원 가중치들에 대한 상한(upper bound)를 정의하고, 전체 가중치들에 대해 절단된 가중치들의 합이 원가중치들의 합과 같아지도록 조정한다. w_{hi} 를 h 층의 i 번째 단위에 대한 최종 가중치라 하고, w_{hB} 를 h 층에 대해 정해진 가중치들의 상한이라 하자. 그러면, h 층의 i 번째 표본단위에 대한 절단 가중치는 다음과 같다.

$$w_{i(T)} = \begin{cases} w_{hi}, & w_{hi} \leq w_{hB} \\ w_{hB}, & w_{hi} \geq w_{hB} \end{cases}$$

전체 표본에 대해 절단된 가중치들의 합이 원 가중치의 합과 같아지도록 조정되어야 한다. 층 내에서 상수인 가중치를 가정하고, F_T 를 원 가중치들의 합과 절단된 가중치들의 합의 비라 하자.

$$F_T = \frac{\sum_h n_h w_h}{\sum_h n_h w_{h(T)}}$$

만일 h 층에 대해 조정된 절단가중치를 $w_{h(T)}^*$ 이라 하면 이는 다음과 같다.

$$w_{h(T)}^* = F_T \times w_{h(T)}$$

이때 $w_{h(T)}^*$ 는 $\sum_h n_h w_{h(T)}^* = \sum_h n_h w_h$ 를 만족한다.

(6) 가중치 조정 시 고려 사항

가중치 조정 시에 고려해야 할 사항을 정리해 보면 다음과 같다.

① 표본설계 시 자체가중설계를 고려하는 것이 바람직하다.

자체가중설계의 장점은 이해가 쉽고 다른 추출설계에 직접적으로 적용이 가능하다는 점이다. 따라서 표본추출설계 시 가능하면 자체가중설계를 고려함으로써 표본설계의 유연성을 확보할 수 있다.

② 자체가중 설계가 어렵다면 가중치 조정단계에 따라 가중값을 부여해야 한다.

만일 자체가중 설계가 표본설계과정에서 고려되지 않았거나 고려하기 어려운 경우라면, 기본가중치와 무응답가중치 그리고 사후가중치 조정단계에 따라 가중치를 계산하여 추정치 산정에 반영해야 할 것이다. 또한 가중치 조정단계에서 이용 가능한 정보를 적극 활용하는 것도 고려해야 한다.

③ 가중치 조정과정에서 이용 가능한 보조정보의 활용을 검토해야 한다.

이용 가능한 정보는 표본과 표본 외부에 모두 존재할 수 있으며, 만일 외부정보를 활용할 수 있다면 이를 적극적으로 활용하는 것이 바람직하다. 사후층화 조정 시에 이러한 외부정보를 사용함으로써 표본과 모집단간의 일치성을 보장할 수 있을 것이다.

④ 가중치를 적용했다면 가중치 효과를 반드시 계산해 보아야 한다.

통상적으로 가중치를 사용하지 않으면, 추정량의 편향이 발생하게 되는 반면에 가중치를 사용하면 추정량의 편향은 줄일 수 있으나, 분산이 증가하는 문제가 있다. 따라서 무조건 가중치를 적용할 것이 아니라 먼저 가중효과를 분산팽창인자를 이용하여 계산해서 적절한 가중치를 적용해야 할 것이다.

⑤ 가중치의 변동을 고려하여 극단 가중치는 추정치에 큰 영향을 줌으로 객관적인 방법으로 절단하여 적용하는 것도 고려할 수 있다.

극단 가중치는 분산을 크게 함으로서 추정치의 정도를 떨어뜨리게 된다.

따라서 적절한 기준 하에서 극단가중치를 조정할 필요가 있다. 이는 총 MSE 차원에서 다루는 것이 합당한데, 왜냐하면 가중치를 조정하여 분산감소의 효과를 볼 수 있는 반면 추정치의 편향이 발생하기 때문이다.

2) 소프트웨어기술자임금실태조사의 가중치 산정

일반적으로 추정단계에서 가중치를 이용하면 모집단에 대한 특성치인 모수에 대한 비편향추정량(unbiased estimator)을 얻을 수 있다. 만약 통계분석 과정에서 가중치를 무시하고 계산한 추정치는 심각한 편향이 발생할 수 있다. 표본의 크기가 큰 대규모 조사에서 문제가 되는 것은 추정량의 편향이기 때문에 추정과정에서 반드시 가중치를 이용해야 한다. 일반적으로 복합표본조사의 가중치는 기본 가중치, 무응답에 대한 조정, 사후층화에 대한 조정 등의 세 가지 요인을 통합하여 산정된다.

이번 표본설계에서는 특성별 평균에 대한 추정치 계산이 가능토록 하였고, 추정량과 추정량의 표준오차공식을 제시하였다.

소프트웨어기술자임금실태조사에서 각종 통계치는 사후층화 조정을 통해서 얻어진 가중치를 이용하여 계산된 것으로 추정량의 분산 계산이 복잡하다. 최근에는 복합표본조사 데이터를 전문적으로 분석할 수 있는 다양한 통계소프트웨어들이 개발되어 널리 활용되고 있다. 대표적인 복합표본조사 분석용 통계소프트웨어는 SAS 8.0판 이상, SUDAAN, WestVarPC, Stata 등이다. 소프트웨어기술자임금실태조사 자료에 대한 고급분석을 위해서는 복합표본조사 분석용 통계소프트웨어를 활용하는 것이 필요하다.

소프트웨어기술자임금실태조사에서 각 기업체에 부여되는 가중치 산출은 총 3단계에 걸쳐 이루어진다. 제 1단계에서는 다음과 같은 방법으로 기본가중치를 계산한다. 종사자 규모(7개 층)와 매출액 규모(6개 층)를 층화변수로 사용하여 총 42개의 층을 고려하고 있으므로, 각 층에 부여되는 기본가중치는 각 층에 속한 모집단 기업체와 표본 기업체의 비율로 표현할 수 있다.

h 층에 속한 표본기업체의 기본가중치는 다음과 같이 표현할 수 있다.

$$\text{기본가중치} : w_{hi}^B = \frac{N_h}{n_h}$$

여기서 N_h 는 h 층의 모집단 기업체 수이고, n_h 는 h 층의 표본 기업체이다.

2단계에서는 다음과 같이 무응답 조정 가중치를 계산한다.

$$\text{무응답 조정 가중치} : w_{hi}^R = \frac{n_h}{r_h}$$

여기서 r_h 는 h 층에서 응답한 기업체의 수이다.

3단계에서는 사후층화를 이용한 가중치를 다음과 같이 산출한다.

$$\text{사후층화가중치} : w_{hi}^F = w_{hi}^B \times w_{hi}^R \times \frac{X_{hi}}{\hat{X}_{hi}}$$

여기서 X_{hi} 는 h 층의 i 번째 기업체의 소프트웨어기술자 총수이며, \hat{X}_{hi} 는 가중합이다.

<표 3-12>에 따라 기본가중치를 부여하고, 최종적으로 가중치를 조정할 때에는 층별로 0인 표본에 대해서는 층을 합쳐 최종 가중치를 산출한다.

<표 3-12> 새로운 종사자 규모별 매출액별 기본 가중치

종사자규모	매출액	모집단	표본기업체	가중치
1~4인	10억 미만	3,661	101	36.2
	10억~50억 미만	370	40	9.3
	50억~100억 미만	12	0	-
	100억~300억 미만	1	0	-
	300억~800억 미만	0	0	-
	800억 이상	0	0	-
5~9인	10억 미만	3,672	125	29.4
	10억~50억 미만	1,685	91	18.5
	50억~100억 미만	89	13	6.8
	100억~300억 미만	15	0	-
	300억~800억 미만	0	0	-
	800억 이상	1	0	-
10~19인	10억 미만	1,531	94	16.3
	10억~50억 미만	2,398	136	17.6
	50억~100억 미만	323	62	5.2
	100억~300억 미만	73	16	4.6
	300억~800억 미만	5	0	-
	800억 이상	0	0	-
20~49인	10억 미만	233	34	6.9
	10억~50억 미만	1,846	139	13.3
	50억~100억 미만	711	92	7.7
	100억~300억 미만	353	55	6.4
	300억~800억 미만	36	0	-
	800억 이상	3	0	-
50~99인	10억 미만	12	7	1.7
	10억~50억 미만	229	50	4.6
	50억~100억 미만	331	64	5.2
	100억~300억 미만	351	62	5.7
	300억~800억 미만	95	15	6.3
	800억 이상	10	0	-
100~299인	10억 미만	1	0	-
	10억~50억 미만	15	0	-
	50억~100억 미만	73	28	2.6
	100억~300억 미만	257	59	4.4
	300억~800억 미만	162	44	3.7
	800억 이상	96	31	3.1
300인 이상	10억 미만	0	0	-
	10억~50억 미만	1	0	-
	50억~100억 미만	7	0	-
	100억~300억 미만	26	26	1.0
	300억~800억 미만	52	41	1.3
	800억 이상	203	75	2.7
계		18,939	1500	12.6

3) 소프트웨어기술자임금실태조사의 모수 추정

(1) 소프트웨어기술자임금의 평균 추정

복합표본조사 데이터를 분석할 때 가중치를 무시하고 분석하면 모수 추정에 심각한 편향(bias)이 발생할 수 있고, 추정량의 분산이 과소평가되어 문제가 된다. 따라서 소프트웨어기술자임금실태조사에서 모집단의 특성치에 대한 추정은 가중치를 이용해야 한다. 만약 단순총계를 사용하면 추정치에 편향이 발생할 수 있다.

소프트웨어기술자임금실태조사에서 모집단 특성치의 추정에 사용될 기호들은 다음과 같다.

$h=1,2,\dots,L$: 층의 수

$i=1,2,\dots,n_h$: h 층 내의 기업체 수

w_{hij} : h 층의 i 번째 기업체의 j 번째 기술자의 가중치

y_{hij} : h 층의 i 번째 기업체의 j 번째 기술자의 기술등급 인력에 대한 측정값

f_h : 추출률

따라서 소프트웨어기술자임금의 평균 추정량 \bar{y} 는 다음과 같다.

$$\bar{y} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}}$$

그리고 소프트웨어기술자임금의 평균 추정량 \bar{y} 의 분산추정량은 다음과 같다,

$$\text{var}(\bar{y}) = \sum_{h=1}^L \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h..})^2$$

여기서 $e_{hi} = w_{hij}(y_{hij} - \bar{y})/w_{...}$, $\bar{e}_{h..} = \left(\sum_{j=1}^{m_{hi}} e_{hij} \right) / m_{hi}$, $w_{...} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$ 이다.

또한 소프트웨어기술자임금의 평균 추정량 \bar{y} 의 상대표준오차는 다음의 식을 통해서 계산한다.

$$\widehat{CV}(\bar{y}) = \frac{\sqrt{\text{var}(\bar{y})}}{\bar{y}} \times 100$$

(2) 각 특성별 소프트웨어기술자임금의 평균 추정

각 특성별 소프트웨어기술자임금의 평균 추정량 \bar{y}_h 는 다음과 같다.

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}}$$

그리고 각 특성별 임금의 평균 추정량 \bar{y}_h 의 분산추정량은 다음과 같다.

$$\text{var}(\bar{y}_h) = \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h..})^2$$

또한 각 특성별 소프트웨어기술자임금의 평균 추정량 \bar{y}_h 의 상대표준오차는 다음의 식을 통해서 계산한다.

$$\widehat{CV}(\bar{y}_h) = \frac{\sqrt{\text{var}(\bar{y}_h)}}{\bar{y}_h} \times 100$$

7. 이상치 검출 및 처리 방안

이 절에서는 이상치 검출 및 처리 방법¹⁾에 대하여 살펴보기로 한다. 기업체 대상의 표본조사에서 이상치의 문제는 그 동안 추정결과의 질을 좌우하는 중요한 문제로 인식되어 수많은 연구가 진행되어 왔다. 특히 기업체 대상의 표본조사에서는 표본추출에 따른 가중치를 포함하고 있고, 연속형 조사변수와 범주형 조사변수가 함께 나타나는 대규모 데이터인 경우가 대부분이다. 또한 기업체 조사에서 가중치는 추정과정에서 중요한 역할을 하게 되고, 이상치를 어떤 방법으로 검출하여 처리하였는가에 따라 통계의 정확도는 중대한 영향을 받게 된다.

일반적으로 표본조사에서의 이상치는 측정이나 조사 과정에서는 아무런 문제없이 정확하게 조사된 값이지만 전체 조사 데이터와 특별하게 동떨어진 극단값이거나 가중치가 커서 그 자료를 제거하거나 포함했을 때 추정에 대단히 큰 영향을 주는 영향점(influential observation)의 의미를 갖는다.

표본조사에 이상치 발생 원인을 정리하면 다음과 같다. 우선 이상치는 모집단의 속성상 오른쪽으로 꼬리가 긴 분포를 갖는 경우(skewed population)에 발생하기 쉽다. 기업체 조사에서는 흔히 주요 관심변수 값이 오른쪽으로 꼬리가 긴 분포를 갖는 경우가 많다. 이와 같이 한쪽으로 치우친 모집단에서 표본조사를 하면 조사값이 다른 값보다 뚜렷이 큰 경우가 발생하고 또는 표본설계에 의한 가중치와 조사값의 곱의 값이 큰 경우가 발생한다.

이상치가 발생한 두 번째 요인으로는 지나치게 큰 가중치로 인해서 발생하는 경우를 들 수 있다. 기업체 조사에서는 일부 표본에 대한 추출확률이 지극히 작은 경우 그 표본에 대한 가중치는 추출확률의 역수가 되므로 매우 큰 값을 가지게 된다.

이상치 발생의 세 번째 원인으로는 비효율적인 표본설계를 들 수 있다. 노후화된 표본추출틀을 이용해서 표본설계를 하는 경우에 추출된 표본은 시간적 변화를 반영하지 못하기 때문에 해당 층의 다른 조사단위와 비교하여 가중치가 대단히 큰 경우가 발생할 수 있다. 이와 같은 비슷한 현상들을 Stratum jumping이라고 하는데, 이 부분에 대한 논의는 Rivest(1999)와 Beaumont and Rivest(2007)를 참고하기 바란다.

이상치 발생의 다른 원인으로는 데이터 생성 과정의 오류(data capture error)를 들 수 있다. 데이터 생성 과정의 오류는 데이터 입력에서 오류가 발생하거나 응답

1) 한국통계학회, 2013

과정에서 오류가 발생하여 매우 크거나 매우 작은 값이 입력되는 경우를 말한다. 이와 같은 이상치는 검출되면 값을 다시 조사하여 수정하거나 분석에서 제외하거나 결측 처리 후 대체방법 등을 이용하여 분석한다.

이상치는 대표성이 없는 이상치(non-representative outlier)와 대표성 있는 이상치(representative outlier)로 구분해 볼 수 있는데, 대표성이 없는 이상치란 값 자체가 잘못된 값이어서 재조사를 통해 자료를 바로 잡거나, 수정을 통해 올바른 값으로 수정할 수 있다. 반면 대표성 있는 이상치는 그 값 자체가 올바른 값이어서 재조사를 통해서도 수정할 수 없는 값이다. 따라서 대표성 있는 이상치는 잘못된 값이 아니고 충분히 모집단에 있을 수 있는 정확한 값이기 때문에 이상치라 하여 무조건 제거하는 것은 재고할 필요가 있다.

1) 이상치 검출 방법

이상치 검출 방법은 한 변수를 사용한 검출 방법과 두 변수간의 관계를 이용한 검출 방법 등으로 구분된다.

(1) 범위 에디팅 방법

범위 에디팅 방법(range edit)은 사용되는 통계량에 따라 두 가지로 구분할 수 있다. 먼저, 한 변수에 대한 이상치를 검출하는 방법으로 해당 조사변수에 대한 표본 평균이 m 이고, 표본표준편차가 s 일 때 다음과 같이 정의된 통계량 d_i 가 사전에 정해진 기준값 C 값에 대하여 $|d_i| > C$ 를 만족할 때 이상치로 검출된다.

$$d_i = \frac{y_i - m}{s}$$

이 방법은 일반 통계 분야에서 사용되는 가장 단순한 방법으로 조사값에 포함된 이상치로 인하여 표본표준편차가 커지게 되어 실제 이상치이지만 이상치로 검출되지 않는 매스킹 효과(masking effect) 현상이 발생할 수 있다. 매스킹 효과에 대한 대처방안으로 로버스트한 방법을 사용하여 중심위치와 산포의 측도를 산출하여 사

용하는 방안을 고려할 수 있다. 예를 들어 표본평균 대신 중앙값을 사용하고, 표본 표준편차대신 사분위수 범위나 $MAD = \text{median}|x - \text{median}(x)|$ 을 사용한다.

(2) 저항 울타리 방법

Thompson(2001)은 이상치를 검출하기 위해 저항 울타리 방법(resistant fences method)을 제안하였다. 두 변수의 비(ratio) 값에 대해서 q_1 이 제 1사분위수이고, q_3 가 제 3사분위수일 때 다음과 같이 정의된 구간 밖에 값이 존재할 때 그 조사값을 이상치로 검출하는 방법이다.

$$(q_1 - k \times (q_3 - q_1), q_3 + k \times (q_3 - q_1))$$

여기서, k 값으로는 1.5, 2, 3 등을 사용한다. 이 방법은 모집단의 분포가 대칭적(symmetric) 분포일 때 주로 사용된다.

(3) 비대칭적 저항 울타리 방법

비대칭적 저항 울타리 방법(Asymmetric resistant fences method)은 모집단의 분포가 비대칭적(skewed) 분포일 때 이상치 검출에 활용할 수 있는 방법이다. Thompson(2001)은 이상치를 검출하기 위해 비대칭적 저항 울타리 방법을 제안하고, 두 변수의 비(ratio) 값에 대하여 다음 범위를 벗어는 경우를 이상치로 검출하였다.

$$(q_1 - k \times (m - q_1), q_3 + k \times (q_3 - m))$$

여기서 m , q_1 , q_3 는 각각 두 변수의 비(ratio)에 대한 중앙값, 제 1사분위수, 제 3사분위수를 나타내고, k 의 값은 3, 4, 6을 사용하였다.

저항 울타리 방법과 비대칭적 저항 울타리 방법은 기본적으로 탐색적 자료분석 기법을 이용하여 이상치를 검출하는 방법으로 한 변수에 대한 적용뿐만 아니라 연관성이 높은 두 변수(예를 들어 전 시점 조사값과 현 시점의 조사값)의 비(ratio)에

대해서 적용함으로써 이상치를 검출할 수 있다.

(4) 로버스트 회귀모형 방법

이 방법은 기본적으로 일반 통계 분야에서 사용되는 최소절사제곱추정량(least trimmed squares estimator : LTS)을 적용하여 로버스트 회귀모형을 적합한 후 잔차값이 $|잔차| \geq 3 \times robust\ MSE$ 이면 이상치로 구분하는 방법이다(Thompson, 2007). 이 방법은 로버스트 회귀모형 이론을 그대로 표본조사 데이터에 적용하여 이상치를 검출하는 방안으로 볼 수 있으며, Hulliger(1999) 등에서와 같이 회귀모형을 기초로 모집단 총계나 평균을 추정하는 경우에 사용되고 있다.

(5) Hidiroglou and Berthelot 방법

Hidiroglou and Berthelot(1986)에서 제시한 이상치 검출 방법은 비(ratio)에 기초한 에디팅 방법으로 구체적인 이상치 검출 과정은 다음과 같다.

$$\textcircled{1} r_i = \frac{x_{i,t+1}}{x_{i,t}}, \text{ 단 } x_{i,t} \text{는 } t \text{시점에서 } i \text{번째 조사단위의 조사값이다.}$$

$$\textcircled{2} s_i = \begin{cases} 1 - \frac{r_{median}}{r_i}, & 0 < r_i < r_{median} \\ \frac{r_i}{r_{median}} - 1, & r_i \geq median \end{cases} \text{ 을 계산한다.}$$

단, r_{median} 은 r_i 의 중앙값이다.

$$\textcircled{3} e_i = s_i \{ \text{MAX}(x_{i,t}, x_{i,t+1}) \}^u \text{ 을 계산한다. 단 } \text{MAX}() \text{는 최대값을 계산하는 함수이고, } u \text{는 자료의 크기 효과를 반영하기 위한 모수로 } 0 \leq u \leq 1 \text{이다. 만약 } u = 0 \text{이면 자료의 크기를 전혀 반영하지 않는 것이며 결과적으로 단지 } e_i \text{는 } s_i \text{에만 의존하게 된다. 반면 } u = 1 \text{인 경우에는 자료의 크기가 } e_i \text{에 매우 큰 영향을 미치게 된다. 실제 자료 분석에서는 } 0.3 \leq u \leq 0.5 \text{가 흔히 사용되고 있다.}$$

- ④ 이상점 판별을 위해 하한과 상한을 이용한 이상점 판별 구간을 정한다. 먼저 e_{median} 을 e_i 의 중앙값이라 하자. 그리고 $d_{q1} = e_{median} - e_{q1}$, $d_{q3} = e_{q3} - e_{median}$, e_{q1} , e_{q3} 를 각각 제 1사분위수와 제 3사분위수라고 하자. 그러면 판별 구간은 다음과 같다.

$$(e_{median} - cd_{q1}, e_{median} + cd_{q3})$$

여기서 c 는 이상점 판별 구간의 길이를 결정하는 모수가 되며 흔히 $c = 1.5$ 를 사용한다. 따라서 이 구간 밖에 있는 자료는 이상점으로 판별된다.

Hidiroglou and Berthelot 방법은 한 조사시점에서 관련성이 깊은 두 변수에 대하여 적용할 수 있는데, 당초 Hidiroglou and Berthelot 방법이 처음 적용되었을 때 두 시점에서의 비(ratio)에 적용했던 것과 같은 방법을 적용할 수 있다.

2) 이상치 처리 방법

데이터 에디팅(data editing) 단계에서 검출된 이상치에 대해서는 우선 조사과정에서 아무런 문제없이 조사되고 정확히 기입된 것인지를 확인해야 한다. 이러한 점검 과정을 통해서 아무런 문제없는 조사값으로 확인된 이상치에 대해서는 추정 과정에서 적절하게 처리될 수 있도록 데이터 상에 표시되어야 하고, 대체(imputation) 과정에서 기증자(donor)로 사용되어서는 안 된다. 일반적으로 검출된 이상치에 대해서는 추정과정에서 이상치의 영향이 감소하도록 값 조정이나 가중치 조정의 과정을 거치게 된다. 이렇게 이상치의 영향을 줄이기 위해서 값 조정(value modification)이나 가중치 조정(weight modification)을 거치게 되면 추정량의 분산은 줄지만 편향(bias)이 발생하게 되기 때문에 평균제곱오차(mean squared error)를 고려하여 이상치의 처리방법을 결정해야 한다.

(1) 가중값 조정(weight modification)

크기가 N 인 모집단으로부터 단순임의추출법에 의해 뽑힌 표본 중 이상치에 해당

하는 부분의 크기와 집합을 각각 n_2 , s_2 라 하고, 이상치가 아닌 부분의 갯수와 집합을 n_1 , s_1 로 정의하자. 여기서 이상치에 해당하는 s_2 에 속한 조사단위의 가중치를 '1'로 낮추어 추정량을 계산하는 방안을 고려할 수 있다. 이 추정량은 이상치에 해당되지 않는 부분의 가중치가 보정되어 모집단의 총계는 다음 추정식에 의해서 추정된다.

$$\hat{Y}_{RI} = \sum_{s_1} \frac{N-n_2}{n_1} y_i + \sum_{s_2} y_i$$

이 추정량의 형태는 s_2 를 전수층으로 구분한 질사표본추출법(cut-off sampling)에 의한 표본에서 구한 모총계 추정량의 형태와 같다. 질사표본추출법은 조사변수 값이 다른 조사단위에 비해서 월등히 클 것으로 예상되는 조사단위들을 하나의 층으로 구성하고, 해당 층에서 전체 조사단위를 표본으로 추출하는 방법을 말한다.

가중값 조정의 다른 방안으로 미국 노동통계청(Bureau of Labor Statistics)의 기업체 조사에서 주로 사용하는 가중값 조정 방법을 살펴본다. BLS 가중값 조정 방법은 4단계로 구분되어 진행된다. 여기서 w_i 는 설계가중치이고, f_i^{REAG} 는 집계 후 보정 요인이며 f_i^{NRF} 는 무응답 보정 요인이다. 또한 f_i^{OAF} 는 이상치 보정 요인이고, f_i^{BMF} 는 벤치마킹 보정 요인이다. 조사값 y_i^* 에 대한 최종가중치는 다음과 같이 표현된다.

$$w_i^{FINAL} = w_i \times f_i^{REAG} \times f_i^{NRF} \times f_i^{OAF} \times f_i^{BMF}$$

모집단 총계에 관련된 최종 추정량은 $\sum_{i=1}^n w_i^{FINAL} y_i^*$ 이다. 최종 가중치에서 이상치 보정 요인 f_i^{OAF} 은 두 부분으로 나누어 보정된다. 먼저 이상치에 해당하는 조사단위의 보정 요인은

$$outlier\ factor = f_i^{OAF} = \frac{1}{w_i}$$

이고, 이상치가 아닌 조사단위의 보정 요인은

$$nonoutlier\ factor = f_i^{OAF} = \frac{\sum_{i=1}^n w_i y_i - \sum_{i \in s_2} y_i}{\sum_{i=1}^n w_i y_i - \sum_{i \in s_2} w_i y_i}$$

이다. y_i 는 표본설계 당시의 조사값이다.

위의 가중치 작성과정에서 이상치 보정 과정은 이상치에 대해서 사실상 설계가중치를 1로 처리하는 것과 같다.

(2) 값 조정(value modification)

이상치 처리 방안의 일종인 값 조정(value modification)은 표본설계에 의해서 결정되는 가중치는 그대로 두고 이상치에 대해 y_i 값을 조정하는 방법이다. 값 조정 방법은 각 조사단위에 대해서 하나의 가중치만 있으면 되기 때문에 편리하다는 장점이 있다.

이상치로 검출된 조사값은 사전에 정해진 원칙에 따라 조정되는 데 일반적으로 널리 사용되고 있는 것은 원저화 방법이다. 이 때, 원저화 방법은 자료를 크기순으로 정렬한 후 작은 값 g 개를 바로 옆의 관측값으로 대체하고 큰 값 g 개를 바로 옆의 관측값으로 대체한 후 계산하는 방법이다.

(3) 가중값과 값의 동시 조정(weight and value modification)

가중값과 값을 동시에 조정하는 방법은 기본적으로 가중값 조정이나 값 조정을 동시에 고려함으로써 이상치의 영향을 줄일 수 있는 방안이다. 일반적으로 표본설계가 이루어진 시점과 실제 조사시점은 시차가 있게 마련이다. 두 시점의 차이가 크게 되면 표본추출틀의 노후화로 인해 같은 층에 속하지만 가중치 편차가 크게 발생하는 경우가 생기게 된다. 특히 문제가 되는 것은 가중치가 지나치게 크게 되는 경우이다. 기업체조사에서는 흔히 가중치와 조사변수의 값의 변동으로 인해 $w_i y_i$ 의 값이 지나치게 커질 수 있다. 그러므로 가중치만 낮추어 조정하거나 또는 값만 조정하는 방법은 이상치를 완전히 조정할 수 없게 된다. 이러한 이유로 제안된 가중

치와 값을 동시에 조정하는 방법은 다음과 같다(Tambay, 1988).

$$w_i y_i^* = \begin{cases} w_i y_i & , K_L < w_i y_i < K_U \\ K_U + \left(y_i - \frac{K_U}{w_i} \right) & , K_U \leq w_i y_i \\ K_L - \left(\frac{K_L}{w_i} - y_i \right) & , w_i y_i \leq K_L \end{cases}$$

앞서 정의된 조정된 가중치와 값을 사용한 총계추정량의 형태는 다음과 같다.

$$\hat{Y}_{W3} = \sum_i w_i y_i^* = \sum_{s_1} w_i y_i + \sum_{s_2} y_i + \sum_{s_2^+} (1 - 1/w_i) K_U + \sum_{s_2^-} (1 - 1/w_i) K_L$$

여기서 K_L 과 K_U 는 각각 가중치가 적용된 조사값에 이상치 검출의 기준값이고, s_2^- , s_2^+ 는 각각 이상치로 검출된 조사단위 중에서 이상치 검출 기준값인 K_L 과 K_U 보다 작거나 큰 경우의 조사단위를 의미한다.

(4) 로버스트 추정(Robust prediction)

모집단의 총계나 평균을 추정하는 경우의 로버스트 추정량은 이상치에 대하여 영향을 덜 받는 추정량을 말한다. 로버스트 추정량을 실제 조사에 활용하기 위해서는 그 형태가 간단하며 통계 이용자에게 쉽게 설명할 수 있어야 한다. 또한 가중치는 조사단위별로 정의되어 여러 관심변수에 대하여 동일한 가중치를 가져야 하고, 추정량의 분산과 편향을 최소화시킬 수 있어야 한다(Hulliger, 1999).

로버스트 추정방법은 앞서 제시한 가중치 조정 방법이나 값 조정 방법과는 달리 가중치나 값에 대한 조정을 하지 않고 이상치의 영향을 덜 받는 모형기반 추정량(model-based estimator)을 이용하는 것이다. 이 방법은 전통적인 설계기반의 추정량을 이용하는 경우에는 거의 사용되지 않지만 스위스 연방통계청에서는 이상치 처리의 중요한 수단으로 사용하고 있다.

3) 소프트웨어기술자임금실태조사의 이상치 검출 및 처리 적용 방안

표본조사에서 무응답 문제와 더불어 이상치(outlier)의 존재는 비표본오차를 증가시키는 또 다른 중요한 요인이다. 이상치를 적절히 처리하지 않고 분석할 경우 모수 추정 결과는 과대 또는 과소 추정될 수 있다. 이러한 과대 또는 과소 추정을 막기 위해서는 이상치를 판별하고 판별된 이상치를 적절히 처리하여야 한다. 소프트웨어기술자임금실태조사에서 이상치가 발생할 경우에는 평균 추정에 미치는 영향이 클 수 있다. 따라서 조사데이터에 대한 이상치 검출과 처리는 추정의 정확도에 많은 영향을 미칠 수 있다.

현재 소프트웨어기술자임금실태조사에서 이상치(outlier)의 기준은 변수의 분포에서 비정상적으로 분포를 벗어난 값으로 정의하고 있으며, 각 변수의 분포에서 비정상적으로 극단값을 갖는 경우나 자료에 타당도가 없는 경우, 비현실적인 변수값들이 해당된다. 하지만 조사된 값이 정확하다면 reasonable outlier로 간주하고 있다.

소프트웨어기술자임금실태조사에서는 전년 값과 비교하여 $\pm 10\%$ 이상 증가 혹은 감소여부 확인하여 이상치로 판단되면 재조사하여 수정하는 방법을 택하고 있다. 특히 응답된 임금 상승률과 전년 응답값을 비교하여 타당여부를 확인하여 이상치로 판단되면 재조사하여 수정 혹은 사유를 기재하고 있다.

소프트웨어기술자임금실태조사의 핵심 변수는 기술자 노임단가라고 할 수 있다. 이들 변수에 대한 이상치 검출 방법으로 비대칭적 저항 울타리 방법이나 Hidiroglou and Berthelot(1986) 방법 등의 적용을 고려해 볼 수 있다. 이 두 방법은 당해 연도 조사결과와 전년도 조사결과의 비(ratio)에 기초한 이상치 검출방법으로 자세한 내용은 앞서 다루었다.

제4장 결론 및 요약

소프트웨어기술자임금실태조사는 1996년 한국소프트웨어산업협회 주관으로 첫 조사를 실시하여 1998년에 통계청으로부터 승인을 받아 매년 실시해 오고 있으며, SW사업을 영위하는 기업체 내 SW기술자의 실지급 임금을 조사하여, 결과를 SW사업 수행시 투입 기술자의 노임단가로 적용하는 중요한 조사이다.

2015년 표본설계 및 조사 데이터에 대한 분석을 통해 개선이 필요한 사항으로 모집단의 정의, 층화 기준(종사자 수 규모, 매출액 규모), 표본크기 산출식, 표본배분 방법, 가중치 산정 및 모수 추정, 이상치 판단 기준 등인 것으로 파악되었다.

이러한 내용을 토대로 본 연구에서는 현행 「소프트웨어기술자임금실태조사」 결과와 표본설계 그리고 최신의 모집단 분석을 통해 소프트웨어기술자의 노임단가를 정확히 공표하기 위한 새로운 표본설계 방안(모집단의 정의, 층화, 표본크기 결정 및 표본배분, 표본추출방법, 가중치 부여 및 모수추정, 이상치 처리 방안 등)을 마련하였다.

본 연구의 주요 결과를 정리하면 다음과 같다.

첫째, 2015년 표본설계 방법을 검토하였고, 소프트웨어기술자임금실태조사 원시자료를 이용해서 주요 문항별 특성을 분석하였다. 그리고 주요 조사항목의 추정결과에 대한 상대표준오차를 계산하여 추정의 정확도를 평가하였다. 소프트웨어기술자 구분별 평균임금을 추정한 결과, 기술사의 평균임금은 8,588,900원이며, 특급 기술자는 7,917,573원, 고급 기술자는 5,734,378원, 중급 기술자는 4,672,278원, 초급 기술자는 3,956,573원 등으로 추정되었다. 한편 기술 등급별 평균 임금에 대한 상대표준오차(CV)를 살펴보면 임금 수준의 변동이 가장 큰 기술등급은 초급 기술자로 10.72%이었으며, 가장 변동이 적은 기술직은 고급기술자로 4.28%로 나타났다. 소프트웨어기술자의 노임단가 등의 주요 변수에 대한 등급별 추정결과의 통계적 정확도는 대체로 안정적인 수준이었고, 부분적으로 표본크기가 작은 등급에서 추정의 정확도가 떨어지는 경우가 부분적으로 나타나고 있음을 알 수 있었다.

둘째, 목표모집단과 조사모집단의 정의를 검토하여, 조사모집단을 ‘한국SW산업협회에 신고한 SW기업체 중 사업유형, 종사자 수 및 매출액 등이 누락된 기업체를

제외한 나머지 18,939개 기업체에 종사하는 소프트웨어기술자 및 DB 입력원'으로 구체적으로 정의하였고, 최신 모집단 자료를 확보한 후 모집단의 특성을 분석하여 새로운 표본설계안에 반영하였다.

셋째, 현행 조사결과와 모집단 분석을 통해 새로운 표본설계에서는 종사자 규모를 7개 층(1~4인, 5~9인, 10~19인, 20~49인, 50~99인, 100~299인, 300인 이상)으로 매출액 규모를 6개 층(10억 미만, 10억~50억 미만, 50억~100억 미만, 100억~300억 미만, 300억~800억 미만, 800억 이상)으로 층화기준을 마련하였다. 그리고 목표 상대 표준오차를 5%로 고려하여 표본크기를 1,500개 이상 수준으로 결정하였다. 소프트웨어기술자 평균 임금 또는 소프트웨어기술자 비율을 고려한 표본배분방법으로 비례배분, 네이만 배분, 멱등배분을 고려하였다. 최종적으로 소프트웨어기술자 비율을 이용한 멱등배분 방법을 통해 표본기업체 규모가 층별로 고르게 배분된 $\alpha=0.4$ 일 때의 멱등배분을 제안하였다. 표본추출방법으로는 종사자 규모와 매출액 규모로 층화한 후 각 층의 기업체를 집락으로 고려하는 추출법을 제시하였다.

넷째, 가중치 조정 방법에 대하여 살펴보았고, 소프트웨어기술자임금실태조사의 기본가중치, 무응답 조정 가중치, 사후층화 가중치를 산출하는 방법에 대하여 다루었으며, 실제 기본가중치를 산출하여 제시하였다.

소프트웨어기술자의 표준임금 산정을 위한 조사로서 기업체 단위의 가중치보다는 소프트웨어기술자 개인별 가중치가 적절하다. 하지만, 소프트웨어기술자의 전체 모집단 규모를 알 수 없기 때문에 기업체 단위의 가중치를 추정과정에 적용해야 하는 한계점이 있다. 이러한 이유로 보다 정확한 소프트웨어기술자의 평균 임금 산출을 위해서는 소프트웨어기술자의 기술 수준별 임금 비중을 고려하여 최종적으로 조정할 필요가 있다.

다섯째, 가중치에 의한 모수 추정 방법으로 소프트웨어기술자임금의 평균 추정량과 분산추정량 및 상대표준오차 식을 제안하였다.

다섯째, 이상치 검출을 위한 여러 가지 방법에 대하여 살펴보았고, 이들 중 이상치 검출 방법으로 당해 연도 조사결과와 전년도 조사결과의 비(ratio)에 기초한 이상치 검출방법인 비대칭적 저항 울타리 방법이나 Hidiroglou and Berthelot(1986) 방

법 등의 적용을 제안하였다.

마지막으로 SW사업 수행시 투입 기술자의 노임단가 파악을 위한 소프트웨어기술자임금실태조사가 더욱 발전하기 위해서는 소프트웨어산업협회에서 지속적으로 모집단 및 표본추출틀을 관리하고 사업유형을 체계화하는 노력이 필요할 것으로 생각된다.

참고문헌

- 대한측량협회(2014). 2014년 측량업체 임금실태 조사보고서.
- 미래창조과학부·소프트웨어정책연구소(2015). SW산업주요통계.
- 미래창조과학부·정보통신산업진흥원(2013). 소프트웨어산업 연간보고서.
- 박홍래(2000). 통계조사론, 영지문화사.
- 통계청(2006). 『소프트웨어기술자임금실태조사』 국가통계 품질진단 연구용역 최종 결과보고서.
- 한국보건사회연구원(2006). 표본추출 및 관리 매뉴얼.
- 한국엔지니어링협회(2015). 2014년도 엔지니어링업체 임금실태조사 보고서.
- 한국조사연구학회(2008). 산업·직업별 고용구조조사 표본설계 연구.
- 한국통계학회(2013). 연구개발활동조사 품질개선 컨설팅 최종결과보고서 - 통계 작성 개선 연구-
- 한국SW산업협회(2015). 2015년 SW기술자 임금실태조사 표본설계내역서.
- Beaumont, J-F. and Rivest, L-P. (2007). A Weight Smoothing Method for Dealing with Stratum Jumpers in Business Surveys. *Proceedings of the Survey Methods Section of SSC Annual Meeting*.
- Bernier, J. and Nobrega, K.(1998). Outlier Detection in Asymmetric Samples: A Comparison of an Inter-quartile Range Method and a Variation of a Sigma Gap Method, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 137-141.
- Bethlehem, J.(2009). *Applied Survey Methods*, Wiley.
- Biemer, P. P. and Lyberg, L. E.(2003). *Introduction to Survey Quality*, John Wiley and Sons.
- Cochran. W. G.(1977). *Sampling Techniques*, John Wiley & Sons, Inc.
- Groves, R. M., Dillman, D. A. Eltinge, J. L., and Little, R. J. A.(2002). *Survey Nonresponse*, John Wiley and Sons.
- Groves, R. M., Fowler, Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R.(2004). *Survey Methodology*, John Wiley and Sons.
- Hansen, M. H. and Hurwitz, W. N.(1946). The Problem of Non-response in Sample Surveys, *Journal of the American Statistical Association*, 41, 517-529.
- Hidiroglou, M. A. and Berthelot, J. M.(1986). Statistical Edit and Imputation for Periodic Surveys, *Survey Methodology*, 12, 73-83.

- Hulliger, B.(1999). Simple and Robust Estimators for Sampling, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 54–63.
- Lavalee, P. and Hidirolou, M. A.(1988). On the Stratification of Skewed Populations. *Survey Methodology*, 14(1), 33–43.
- Rivest, L.–P.(1999). Stratum Jumpers: Can We Avoid Them? *Proceedings of the Survey Research Methods Section, American Statistical Association*, 64–72.
- Rouseeuw and Zomeren(1990). Unmasking Multivariate Outliers and Leverage Points, *Journal of the American Statistical Association*, 85, 633–639.
- SAS/STAT User's Guide, Version 8, SAS Publishing.
- Srinath, K. P. and Carpenter, R. M.(1995). Sampling Methods for Repeated Business Surveys, *Business Survey Methods*, Eds. Cox, B.G. et al., Wiley.
- Thompson, K.(2007). Investigation of Macro Editing Techniques for Outlier Detection in Survey Data, *Proceedings of the Third International Conference on Establishment Surveys, American Statistical Association*.
- Thompson, K. J.(2001). Ratio Edit Tolerance Development Using Variations of Exploratory Data Analysis(EDA) Resistant Fences Methods, *Statistical Policy Working Paper 29, a Federal Committee on Statistical Methodology Conference Paper*.

부록 1. 소프트웨어기술자 등급분류 기준

소프트웨어 기술자 등급분류 기준표 - 2009년 8월 1일 이후

구분	기술자격자	학력·경력자
기술사	· 기술사	
특급 기술자	· 고급기술자 자격 취득 후 3년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람	
고급 기술자	· 중급기술자 자격 취득 후 3년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 박사학위를 가진 자로서 기사자격을 취득한 자	
중급 기술자	· 기사 자격을 취득한 자로서 3년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 산업기사 자격을 취득한 자로서 7년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 기사자격을 취득한 자로서 석사학위 취득 후 2년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람	
초급 기술자	· 기사 자격을 취득한 자 · 산업기사 이상의 자격을 취득한 자	· 전문학사 이상의 학위를 가진 자 · 고등학교를 졸업한 후 3년 이상 소프트웨어 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람
고급 기능사	· 산업기사의 자격을 취득한 자로서 4년 이상 소프트웨어 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 기능사의 자격을 취득한 자로서 7년 이상 소프트웨어 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람	
중급 기능사	· 산업기사의 자격을 취득한 자 · 기능사의 자격을 취득한 자로서 3년 이상 소프트웨어 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람	
초급 기능사	· 기능사의 자격을 취득한 자	

소프트웨어 기술자 등급분류 기준표 - 2009년 7월 31일 이전

구분	기술자격 및 경험 기준	학력 및 경험기준
기술사	· 기술사	
특급 기술자	· 기사자격을 가진 자로서 10년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 산업기사자격을 가진 자로서 13년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람	· 박사학위를 가진 자로서 3년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 석사학위를 가진 자로서 9년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 학사학위를 가진 자로서 12년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 전문대학을 졸업한 자로서 15년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람
고급 기술자	· 기사자격을 가진 자로서 7년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 산업기사자격을 가진 자로서 10년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람	· 박사학위를 가진 자 · 석사학위를 가진 자로서 6년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 학사학위를 가진 자로서 9년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 전문대학을 졸업한 자로서 12년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 고등학교를 졸업한 자로서 15년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람
중급 기술자	· 기사자격을 가진 자로서 4년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 산업기사자격을 가진 자로서 7년 이상 해당 기술 분야에서 일정기간 경력을 갖추거나 근무한 사람	· 석사학위를 가진 자로서 3년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 학사학위를 가진 자로서 6년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 전문대학을 졸업한 자로서 9년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 고등학교를 졸업한 자로서 12년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람
초급 기술자	· 기사자격을 가진 자 · 산업기사자격을 가진 자	· 석사학위를 가진 자 · 학사학위를 가진 자 · 전문대학을 졸업한 자 · 고등학교를 졸업한 자로서 3년 이상 해당기술 분야에서 일정기간 경력을 갖추거나 근무한 사람
고급 기능사	· 산업기사자격을 가진 자로서 4년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 기능사자격을 가진 자로서 7년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람	· 기능대학 또는 전문대학을 졸업한 자로서 4년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 고등학교를 졸업한 자로서 7년 이상 해당기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 직업훈련기관의 교육을 이수한 자로서 7년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 기능실기시험을 합격한 자로서 10년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람
중급 기능사	· 산업기사자격을 가진 자 · 기능사자격을 가진 자로서 3년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람	· 기능대학 또는 전문대학을 졸업한 자 · 고등학교를 졸업한 자로서 3년 이상 해당기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 직업훈련기관의 교육을 이수한 자로서 5년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 기능실기시험을 합격한 자로서 5년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람 · 그 밖에 10년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람
초급 기능사	· 기능사자격을 가진 자	· 고등학교를 졸업한 자 · 직업훈련기관의 교육을 이수한 자 · 기능실기시험을 합격한 자 · 그 밖에 5년 이상 해당 기능 분야에서 일정기간 경력을 갖추거나 근무한 사람

※ “소프트웨어 기술자”의 범위는 소프트웨어산업진흥법 제2조 5항에서 정한 바에 따름

소프트웨어산업 진흥법 제2조(정의)

⑤ "소프트웨어기술자"란 「국가기술자격법」에 따라 정보처리 분야의 기술자격을 취득한 사람 또는 소프트웨어 기술 분야에서 대통령령으로 정하는 학력이나 경력을 가진 사람을 말한다.

소프트웨어산업 진흥법 시행령 제1조의2(소프트웨어기술자의 범위)

① 「소프트웨어산업 진흥법」(이하 "법"이라 한다) 제2조제5호에서 "대통령령으로 정하는 학력이나 경력을 가진 사람"이란 다음 각 호의 어느 하나에 해당하는 사람을 말한다.

1. 「초·중등교육법」 제2조제3호 또는 「고등교육법」 제2조에 따른 각급 학교에서 소프트웨어 기술 분야를 전공한 사람
 2. 소프트웨어 기술을 가진 사람으로서 소프트웨어 기술 분야에서 일정 기간 경력을 갖추거나 근무한 사람
 3. 그 밖에 제1호 및 제2호에 따른 기준과 같거나 그 이상의 학력 또는 경력이 있다고 인정되는 사람
- ② 제1항에 따른 소프트웨어기술자에 대한 세부적인 인정 기준 및 절차·방법 등은 미래창조과학부장관이 정하여 고시한다.

* “기술자격자” 중 기술사, 기사, 산업기사, 기능사는 「국가기술자격법」의 기술자격종목 중 다음 각 목의 정보처리 분야 기술자격을 취득한 자를 말한다.

가. 기술사: 정보관리, 컴퓨터시스템응용

나. 기사: 정보처리, 전자계산기조직응용, 정보보안

다. 산업기사: 정보처리, 사무자동화, 정보보안

라. 기능사: 정보처리, 정보기기응용

부록 2. 소프트웨어기술자임금실태조사 조사표

2015년 소프트웨어기술자 임금실태 조사표

(작성기관 : 한국소프트웨어산업협회)

제37501호

《 조사의 목적 》

본 SW기술자 임금실태 조사는 통계청의 승인을 받아 소프트웨어기술자 임금현황을 파악하여 이를 통한 정부의 정책 지원 방향의 기반을 마련하기 위한 조사로 실지급 임금이 정확히 파악될 수 있도록 빠짐없이 작성하여 주시기 바랍니다.

《 조사의 법적근거 》

통계법 제18조 (통계작성의 승인)

소프트웨어산업진흥법 제22조(소프트웨어사업의 대가지급)

소프트웨어산업진흥법 시행령 제16조(소프트웨어기술자의 노임단가)

국가를 당사자로 하는 계약에 관한 법률 시행규칙 제7조(원가계산을 할 때 단위당 가격의 기준)

《 조사대상 및 방법 》

소프트웨어사업자 신고업체 및 한국소프트웨어산업협회 회원사를 대상으로 E-Mail, Fax조사 및 현장방문조사 병행

《 작성 및 제출 》

본 조사는 소프트웨어기술자의 월평균 임금(연봉/12)을 기준으로 작성하시어 2015년 7월 17일(금)까지 이메일(agnes@sw.or.kr)로 송부 부탁드립니다.

《 문의처 》

한국소프트웨어산업협회 담당자 : 이경숙 책임(02-2188-6931)

※ 본 조사표는 통계법 제33조에 의거 개별업체의 조사내용은 엄격히 비밀이 보장되며 통계자료 목적 이외에는 사용되지 않습니다.

3. 인력현황(2015년 월평균 기준으로 작성)

1) 인력구성

총인원	명		①SW사업전담 기술인력	명		SW사업관련 사무직 및 기타	명	
	남	여		남	여		남	여
	명	명		명	명		명	명

※ 총인원은 소프트웨어 사업이 전업이 아닌 경우 타 사업분야의 인력까지 포함한 전체종사자 수

2) 소프트웨어 전담 기술인력 현황

등급 구분	기술사		특급기술자		고급기술자		중급기술자		초급기술자		고급기능사		중급기능사		초급기능사		DB입력원	
	SW전담 기술인력	명		명		명		명		명		명		명		명		명
남		여	남	여	남	여	남	여	남	여	남	여	남	여	남	여	남	여
명		명	명	명	명	명	명	명	명	명	명	명	명	명	명	명	명	명
															명			
															② 합계		남	여

※ 등급분류 기준 : 첨부작성매뉴얼 참고

※ 인력구성의 ①항과 2)소프트웨어 전담 기술인력 현황의 ②항은 일치하여야 함

4. SW기술자 등급별 임금현황

(▶소프트웨어 전담 기술인력이 있는 경우만 기재)

(단위 : 명, 원)

구 분	① 기술사					② 특급기술자					③ 고급기술자					④ 중급기술자																								
	만	천	백	십	일	만	천	백	십	일	만	천	백	십	일	만	천	백	십	일																				
대상인원																																								
단위(원)	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일
① 기본급여																																								
② 제수당																																								
③ 월간상여금																																								
④ 월간퇴직 급여충당금																																								
⑤ 법인부담금																																								
합계(①~⑤)																																								
평균 연봉인상 률(%)																																								
구 분	⑤ 초급기술사					⑥ 고급기능사					⑦ 중급기능사					⑧ 초급기능사																								
단위(명)	만	천	백	십	일	만	천	백	십	일	만	천	백	십	일	만	천	백	십	일																				
대상인원																																								
단위(원)	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일	십	억	천	백	십	만	천	백	십	일
① 기본급여																																								
② 제수당																																								
③ 월간상여금																																								
④ 월간퇴직 급여충당금																																								
⑤ 법인부담금																																								
합계(①~⑤)																																								
평균 연봉인상 률(%)																																								
구 분	⑨ DB입력원																																							
단위(명)	만	천	백	십	일																																			
대상인원																																								
단위(원)	십	억	천	백	십																					만	천	백	십	일										
① 기본급여																																								
② 제수당																																								
③ 월간상여금																																								
④ 월간퇴직 급여충당금																																								
⑤ 법인부담금																																								
합계(①~⑤)																																								
평균 연봉인상 률(%)																																								

5. 임금동향

▣ 아래 3가지 사항 중 해당되는 항목에 한해서 변동율과 그 요인을 기재하여 주시기 바랍니다.

전년대비 임금이 상승한 경우 임금상승요인 (가장 영향이 큰 항목 한가지 "O"표)	① 기술인력 부족 () ② 다른업체와의 보조 () ③ 물가상승 () ④ 경영수지 향상 () ⑤ 노사관계 안정 필요 () ⑥ 경기호전 () ⑦ 기타()
전년대비 임금이 인화된 경우임금인하요인 (가장 영향이 큰 항목 한가지 "O"표)	① 기술인력 공급증가 () ② 경영수지악화 () ③ 물가하락 () ④ 경기악화 () ⑤ 기타()
전년대비 임금이 동일한 경우	① 노동력 수급안정 () ② 물가안정 () ③ 기타()

6. 기타의견

▣ 2015년 소프트웨어기술자 임금실태 조사와 관련한 귀하의 의견을 자유롭게 기재하여 주시기 바랍니다.

★ 작성해 주셔서 대단히 감사합니다 ★

부록 3. 개정된 중소기업기본법

「중소기업기본법」(이하 "법"이라 한다) 제2조 제1항 제1호에 따른 중소기업은 다음 각 호의 기준을 모두 갖춘 기업으로 한다. <개정 2014.4.14>

1. 다음 각 목의 요건을 모두 갖춘 기업일 것

- 가. 해당 기업이 영위하는 주된 업종과 해당 기업의 평균매출액 또는 연간매출액(이하 "평균매출액등"이라 한다)이 별표 1의 기준에 맞을 것 (공평성을 기하기 위해 매출기준액은 업종마다 다르다. 아래 표에 정리되어 있다)
- 나. 자산총액이 5천억원 미만일 것

2. 소유와 경영의 실질적인 독립성이 다음 각 목의 어느 하나에 해당하지 아니하는 기업일 것

- 가. 「독점규제 및 공정거래에 관한 법률」 제14조제1항에 따른 상호출자제한기업집단 또는 채무보증제한기업집단에 속하는 회사
- 나. 자산총액이 5천억원 이상인 법인(외국법인을 포함하되, 제3조의2제3항 각 호의 어느 하나에 해당하는 자는 제외한다)이 주식등의 100분의 30 이상을 직접적 또는 간접적으로 소유한 경우로서 최다출자자인 기업. 이 경우 최다출자자는 해당 기업의 주식 등을 소유한 법인 또는 개인으로서 단독으로 또는 다음의 어느 하나에 해당하는 자와 합산하여 해당 기업의 주식등을 가장 많이 소유한 자를 말하며, 주식 등의 간접소유 비율에 관하여는 「국제조세조정에 관한 법률 시행령」 제2조제2항을 준용한다.

1) 주식등을 소유한 자가 법인인 경우: 그 법인의 임원

2) 주식등을 소유한 자가 1)에 해당하지 아니하는 개인인 경우: 그 개인의 친족

- 다. 관계기업에 속하는 기업의 경우에는 제7조의4에 따라 산정한 평균매출액등이 별표 1의 기준에 맞지 아니하는 기업② 법 제2조제1항제2호에서 "대통령령으로 정하는 사회적기업"이란 영리를 주된 목적으로 하지 아니하는 사회적기업으로서 다음 각 호의 기준을 모두 갖춘 기업으로 한다. <개정 2014.4.14>

별표1.

주된 업종별 평균매출액등의 규모 기준(제3조 제1항 제1호 가목 관련)		
해당 기업의 주된 업종	분류 기호	규모 기준
1. 의복, 의복액세서리 및 모피제품 제조업 2. 가죽, 가방 및 신발 제조업 3. 펄프, 종이 및 종이제품 제조업 4. 1차 금속 제조업 5. 전기장비 제조업 6. 가구 제조업	C14 C15 C17 C24 C28 C32	평균매출액등 1,500억원 이 하
7. 농업, 임업 및 어업 8. 광업 9. 식료품 제조업 10. 담배 제조업 11. 섬유제품 제조업(의복 제조업은 제외한다) 12. 목재 및 나무제품 제조업(가구 제조업은 제외한다) 13. 코크스, 연탄 및 석유정제품 제조업 14. 화학물질 및 화학제품 제조업(의약품 제조업은 제외한다) 15. 고무제품 및 플라스틱제품 제조업 16. 금속가공제품 제조업(기계 및 가구 제조업은 제외한다) 17. 전자부품, 컴퓨터, 영상, 음향 및 통신장비 제조업 18. 그 밖의 기계 및 장비 제조업 19. 자동차 및 트레일러 제조업 20. 그 밖의 운송장비 제조업 21. 전기, 가스, 증기 및 수도사업 22. 건설업 23. 도매 및 소매업	A B C10 C12 C13 C16 C19 C20 C22 C25 C26 C29 C30 C31 D F G	평균매출액등 1,000억원 이 하
24. 음료 제조업 25. 인쇄 및 기록매체 복제업 26. 의료용 물질 및 의약품 제조업 27. 비금속 광물제품 제조업 28. 의료, 정밀, 광학기기 및 시계 제조업 29. 그 밖의 제품 제조업 30. 하수·폐기물 처리, 원료재생 및 환경복원업 31. 운수업 32. 출판, 영상, 방송통신 및 정보서비스업	C11 C18 C21 C23 C27 C33 E H J	평균매출액등 800억원 이하
33. 전문, 과학 및 기술 서비스업 34. 사업시설관리 및 사업지원 서비스업 35. 보건업 및 사회복지 서비스업 36. 예술, 스포츠 및 여가 관련 서비스업 37. 수리(修理) 및 기타 개인 서비스업	M N Q R S	평균매출액등 600억원 이하
38. 숙박 및 음식점업 39. 금융 및 보험업 40. 부동산업 및 임대업 41. 교육 서비스업	I K L P	평균매출액등 400억원 이하

비고 : 해당 기업의 주된 업종의 분류 및 분류기호는 「통계법」 제22조에 따라 통계청장이 고시한 한국표준산업분류에 따른다.

부록 4. 표본추출 및 모수추정을 위한 SAS 프로그램

```
/*표본추출을 위한 SAS 프로그램(n=1510)*/
proc freq data=s.pop2015_2;/*층화 변수를 지정*/
tables emp*size/nopercent norow nocol out=freq;
run;
data freq;
set freq;
chg=_n_;
run;
/*층화변수 chg를 모집단 파일에 머지*/
proc sort data=s.pop2015_2;by emp size;run;
proc sort data=freq; by emp size;run;
data s.pop2015_2;
merge s.pop2015_2 freq;
by emp size;
run;
proc sort data=s.pop2015_2;by chg;run;/*층화변수기준으로 오름차순 정렬*/
/*표본수 1510개 기업체를 층화계통 추출*/
proc surveyselect data=s.pop2015_2 m=sys seed=1234567 out=sam1500
n=(101 40 1 1
125 91 13 1 1
94 136 62 16
34 139 92 55 1
7 50 64 62 15 1
1 1 28 59 44 31
1 1 1 26 41 75);
strata chg;
run;

/*모수 추정을 위한 프로그램(기술자 1인당 평균 임금의 추정)*/
/*기술자별 평균임금=기술자별 총임금/기술자별 총인원수*/
data s.sam2015_1;/*임금 총액 산출(평균 임금*기술자수)*/
set s.sam2015_1;
x1=_col20*_col2;
x2=_col21*_col3;
```

```

x3=_col22*_col4;
x4=_col23*_col5;
x5=_col24*_col6;
x6=_col25*_col7;
x7=_col26*_col8;
x8=_col27*_col9;
x9=_col28*_col10;
run;
/*모집단과 표본 데이터의 결합*/
proc sort data=s.pop2015_2;by id;run;
proc sort data=s.sam2015_1;by id;run;
data s.sam2015_1;
set s.sam2015_1; s=1; run;
data s.sam_imsi;
merge s.pop2015_2 s.sam2015_1;
by id;
run;
data s.sam_imsi; /*표본데이터의 총임금 산출*/
set s.sam_imsi;
if s=1;
tot_income=sum(_col20,_col21,_col22,_col23,_col24,_col25,_col26,_col27,_col28);
run;
proc means data=s.sam_imsi mean std cv maxdec=1
var tot_income;
*class emp size;
run;
proc sort data=s.sam_imsi; by emp size;run
proc surveymeans data=s.sam_imsi mean stderr cv ;
var tot_income;
*strata emp size;
domain emp*size;
run;
proc means data=s.sam_imsi n sum maxdec=0
var _col2 - _col10;
class emp size;
run;
proc freq data=s.sam_imsi;
tables _COL29 -_COL37;

```

```
run;
```

```
proc surveymeans data=s.sam_imsi_1 ratio stderr;  
var x1-x9 _col2-_col10;/*x1-x9: 기술자 총임금, _col2-_col10: 기술자 수*/  
ratio x1 /_col2;  
ratio x2/_col3;  
ratio x3/_col4;  
ratio x4/_col5;  
ratio x5/_col6;  
ratio x6/_col7;  
ratio x7/_col8;  
ratio x8/_col9;  
ratio x9/_col10;  
strata emp1 size1;/*종사자 수 및 매출액 규모별 층화*/  
domain emp1*size1;/*층별 추정*/  
weight wt;  
cluster id;/*집락을 기업체로 설정*/  
run;
```