

2011년도 국가통계 품질개선지원 연구용역

『엔지니어링업체 임금실태조사』
품질개선지원 최종결과보고서
- 전수조사에서 표본조사로의 전환 방안 -

2011. 9.

제 출 문

제 출 문

통계청장 귀하

본 보고서를 『엔지니어링업체 임금실태조사 통계』
품질개선지원 연구용역 과제의 최종 연구결과물로 제출
합니다.

[개선지원 부문]: 전수조사에서 표본조사로의 전환 방안

2011년 9월

한국통계진흥원장 이 동 명 ㉠

연구진

책임연구원	임 명 선 (한국통계진흥원 사무국장)
연구원	임 대 철 (한국통계진흥원 조사부 과장)
연구원	김 중 환 (한국통계진흥원 기획연구부 과장)

최종결과보고서 요약문

연구과제명	「엔지니어링업체 임금실태조사」 전수조사에서 표본조사로의 전환 방안
주제어	엔지니어링, 임금실태조사
연구기간	2011. 06. 17 ~ 2011. 9. 30
연구기관	(재)한국통계진흥원
연구진	임명선, 임대철, 김종환
<p>본 연구는 한국엔지니어링협회에서 매년 조사하는 엔지니어링업체 임금실태조사의 표본설계 개선을 위한 연구이다. 지난 2007년 실시한 통계청의 정기품질진단 결과에서 ‘표본설계’ 부분에 대해 제기된 문제점에 대해 개선안을 제공하는데 그 목적이 있다.</p> <ul style="list-style-type: none"> ○ 2010년 엔지니어링업체 임금실태조사에 대한 특성을 살펴보고 표본조사로 전환이 가능한지를 검토하였다. ○ 검토결과, 엔지니어링협회에 신고된 자료를 검토하여 조사방법을 표본조사 전환할 수 있도록 방법론을 제시하였다. <ul style="list-style-type: none"> - 먼저 신고된 자료를 토대로 모집단을 정의하였고 표본설계에 활용가능한 표본추출틀을 마련하였다. - 업종 중 모집단 규모가 작은 업종은 전수조사를 그대로 유지하도록 하고 나머지 업종에 대해서는 업종별 표본규모를 제시하였다. - 각 층의 표본이 적절하게 추출되도록 종사자수를 정렬하여 추출하는 계통추출방법을 제시하였다. - 추출된 표본에 대해 조사완료 후 추정이 가능하도록 가중치 산정방법을 제시하고 모집단에 대한 추정식을 제시하였다. ○ 본 연구를 통해 조사방법을 전수조사에서 표본조사로의 전환함에 따라 정확성과 신뢰성에 대해 불확실성이 사라질 것으로 예상된다. 	

목 차

제1장 서론	1
1절 연구배경	1
2절 연구목적과 필요성	1
(1) 필요성	1
(2) 목적	1
3절 연구내용 및 방법	3
(1) 연구방법	3
(2) 모집단 분석	3
(3) 표본설계 방안제시	4
4절 연구흐름도 및 협의내용	4
(1) 연구흐름도	4
(2) 연구협의 내용	6
제2장 통계 및 개선과제 개요	8
1절 통계개요	8
(1) 조사목적	8
(2) 조사설계	8
(3) 조사현황	9
2절 개선과제	10
(1) 현황	10
(2) 문제점	10
(3) 개선사항	11
3절 유사조사 분석	11

제3장 표본설계	13
1절 모집단 분석	13
(1) 모집단 설정	13
(2) 층화	17
(3) 조사모집단 현황	18
2절 표본설계	19
(1) 표본크기	19
(2) 표본추출	22
(3) 모의실험	23
3절 추정	24
(1) 가중값	24
(2) 추정	25
제4장 결론 및 제안	27

- 표 목 차 -

<표 1> 연구내용 및 방법	3
<표 2> 회수 및 집계 현황	9
<표 3> 산업별·종사자규모별 모집단 현황	13
<표 4> 표본추출틀 검토사항	15
<표 5> 목표모집단 현황	16
<표 6> 업종별 사업체수 및 전수/표본 구분	17
<표 7> 업종별·종사자규모별 구분기준	18
<표 8> 목표모집단 현황	18
<표 9> 목차오차에 따른 표본수	20
<표 10> 표본수(목표오차 5.0%)	21

- 그림 목 차 -

<그림 1> 연구흐름도	5
--------------------	---

제1장 서론

1절 연구배경

국가통계 품질진단 결과로 조기 품질개선이 필요한 통계에 대하여 개선안과 구체적인 실행방안을 제시함으로써 해당 통계의 신속한 품질개선 및 정책 자료로의 활용도를 제고하기 위하여 국가통계정기품질진단의 내용을 토대로 표본설계 부분에서 이론적·실증적 연구가 필요하다.

2절 연구목적과 필요성

(1) 필요성

일반적으로 통계조사는 모집단의 일부분인 표본(sample)에 의존한다. 조사대상 전체의 집합인 모집단(population)을 대상으로 하는 통계조사는 현실적으로 거의 불가능하기 때문이다. 따라서 표본은 모집단을 가장 잘 대표할 수 있어야 한다.

표본추출을 어떻게 하는가에 따라서 통계조사의 성패를 좌우할 뿐만 아니라 나아가 통계의 품질에도 직접적인 영향을 미치게 된다. 표본추출을 하기 위해서는 표본의 크기(규모)나 추출방법 등을 결정하는 일련의 과정이 선행되어야 하는데 이를 표본설계라 하며 실제 모집단과의 차이가 최소가 되도록 하는 것이 가장 중요하다.

(2) 목적

통계청의 2007년도 정기품질진단 결과에서 '표본설계' 부분에 대해 제기된 문제점과 개선안은 다음과 같다.

□ 문제점

○ 조사 회수율

- 전수조사임에도 불구하고 조사 회수율이 80% 정도로 나타나고 있으며 나머지 20% 정도는 영세규모 업체로서 조사회수율이 낮은 편이며 업종별 회수율의 편차가 높게 나타난다.

○ 사후가중치 미적용

- 전수조사이므로 표본가중치는 필요가 없으나 조사 회수율이 100%가 되지 않는 점을 고려할 때 조사된 자료에 대해 사후가중치 적용이나 보정 등이 시행되지 않고 있다.

○ 조사 방법 구분

- 기본적인 조사방법은 우편조사를 근간으로 하고 Fax를 이용한 조사가 부수적으로 사용됨에 따라 조사방법으로 인한 불성실한 응답뿐만 아니라 조사표를 이해 못하는 경우 비표본 오차를 발생 가능성이 높은 편이다.

□ 개선안

- 방문면접조사를 수행할 경우 조사비용에 대한 한계점이 있으므로 전수조사보다는 표본조사로 전환 검토가 필요하다.
- 모집단(Population)에 대한 정확한 정의가 필요하고 표본설계시 사용가능한 표본추출틀(Sampling Frame)의 확정이 필요하다.
- 모집단을 대표할 수 있는 최적의 표본설계 방법을 연구한다.
- 모수에 대한 추정이 가능하도록 추정식을 제시한다.

2007년도에 지적된 문제점들과 개선사항들이 현재까지 반영되지 않고

있기 때문에 본 연구를 통해 앞서 제기된 문제점을 보완하고 요구되는 개선안을 최대한 반영한 최적의 표본설계방법을 찾고자 한다.

3절 연구내용 및 방법

(1) 연구방법

- 「엔지니어링임금실태조사」의 연구내용 및 방법은 아래와 같다.

<표 1> 연구내용 및 방법

연구 내용	연구 방법
- 엔지니어링임금실태조사의 조사설계, 조사방법 등 검토	- 통계작성기관 방문 면담 - 통계청 '국가통계품질진단'보고서 검토 - 유사 조사 수집 및 검토
- 조사 모집단 검토 및 문제점 파악	- 통계작성기관으로부터 제공받은 조사모집단에 대해 분류 기준 및 표본설계 활용시 문제점 파악
- 표본추출틀 구축	- 표본설계시 사용가능한 표본추출틀 정의
- 적합한 표본설계 제시	- 통계작성기관과의 협의와 활용방안을 고려한 적합한 표본설계 방법을 제시
- 최종 표본설계 방안 정립	- 통계작성기관과의 협의회와 자문회의를 통한 최종 실행가능한 표본설계방안 제시

(2) 모집단 분석

- 목표모집단
 - 엔지니어링기술진흥법에 의거하여 신고한 엔지니어링 활동주체
- 조사모집단
 - 엔지니어링기술진흥법 제21조에 의거하여 한국엔지니어링협회에

신고한 사업체

○ 검토 사항

- 현 상황에서 가장 적절한 모집단을 결정한다.
- 신고된 협회자료를 검토하여 표본추출틀을 결정한다.

(3) 표본설계 방안제시

- 모집단 현황을 분석하여 사용가능한 표본추출틀을 정의한다.
- 목표모집단에 가장 적합한 표본 추출방법을 제시하여 현실가능한 표본설계 방안을 제시한다.

4절 연구흐름도 및 협의내용

(1) 연구흐름도

- 본 연구의 전체 흐름도는 다음과 같은 과정을 통해 이루어진다.
- 먼저 조사대상이 되는 모집단인 조사모집단을 명확하게 정의하고 정확한 표본조사를 위해서는 개념적으로 정의한 목표모집단과 표본추출을 위한 실제 조사대상인 조사모집단의 차이가 최소가 되도록 하는 것이 중요하다.
- 조사모집단이 정해지면 조사모집단을 최대한 포함하는 표본추출틀을 준비하고 조사목적에 맞으면서 이용 가능한 표본추출틀을 확보하여 표본설계시 사용가능한 항목들을 포함하여 결정한다.
- 또한 조사모집단의 조사사업체가 빠지거나 중복되는 경우가 없도록 해야 하며, 조사모집단은 가능한 한 동일한 표본추출틀을 사용해서 일관성을 유지하도록 한다.

- 표본추출틀이 마련되면, 조사일정을 수립하고 일정에 따른 비용을 산정하여 예산을 배정한다.
- 그런 다음, 실제 조사에 필요한 표본크기와 가장 효율적인 추출법을 정하고, 그에 따라 층화와 표본배분, 표본추출단계, 조사방법 등을 결정하는 등 구체적인 계획들을 마련한다.
- 마지막으로 모수 추정식을 제시한다.



<그림 1> 연구흐름도

(2) 연구협의 내용

□ 협의 일시 및 장소

- 일 시 : 2011. 8. 5(금)
- 장 소 : 한국엔지니어링협회
- 참석자 : 황창준(한국엔지니어링협회 정보·통계팀 과장), 김효정(한국엔지니어링협회 정보·통계팀 사원), 김용환(한국통계진흥원 조사부 부장), 임대철(한국통계진흥원 조사부 과장)

□ 협의내용

- 모집단 정의
 - 조사모집단은 엔지니어링산업진흥법 제21조에 의하여 신고된 엔지니어링사업체로서 1994년도에는 761개사에서 2010년 12월말에는 4,592개사로 급격하게 늘어나고 있는 추세이다.
 - 신고된 사업체에 대해서 종사자수, 매출액 등의 항목을 기입하도록 되어있으며 조사에서는 종사자수와 매출액등의 항목에 대해서는 조사하지 않았다.
- 모집단 층화
 - 업종에 따라서 기계, 선박, 항공·우주, 금속, 전기·전자, 통신정보처리, 화학, 관업자원, 건설, 환경, 농림, 해양수산, 산업관리, 응용이학으로 구분할 수 있다.
 - 또한, 엔지니어링활동에 따라서 사업부문을 원자력발전, 산업공장, 건설, 기타 엔지니어링활동으로 구분할 수 있다.

○ 조사의 문제점

- 매년 우편조사를 통해 조사를 실시하였고 회수율은 80% 정도로 높게 나타났다. 계약직을 고용하여 약 5개월 정도 조사를 실시하며 우편이 회수되지 않는 사업체에 대해서는 독려전화를 통해 조사를 유도하고 있다.
- 하지만 조사방법이 우편조사다 보니 조사기간이 길어지고 영세 사업체에서의 조사회수율이 낮은 편이고 협조가 열악하다.

○ 표본설계에 사용할 자료

- 2011년 6월말 기준 한국엔지니어링협회에 신고한 엔지니어링활동 사업체 자료를 사용한다.

제2장 통계 및 개선과제 개요

1절 통계개요

(1) 조사목적

엔지니어링활동을 하는 사업체에 종사하는 기술자의 실지급 임금수준을 조사하여, 그 결과를 엔지니어링사업 수행시 기술자의 노임단가로 적용할 수 있도록 해당업체 및 유관기관에 제공하고, 엔지니어링사업체에 종사하는 기술자의 임금동향을 파악하기 위한 기초자료로 활용한다.

(2) 조사설계

- 조사 대상
 - 엔지니어링산업 진흥법 제21조에 따라 신고한 엔지니어링 사업자
- 조사 항목
 - 기본급, 제수당, 월간상여금, 월간퇴직급여충당금, 월간사회보험료 (회사부담분)
- 조사 방법
 - 우편조사 또는 FAX, 인터넷 조사
- 조사 기준
 - 조사 임금 : 각 년도의 만근한 기술자의 실지급 임금
 - 조사 기간 : 8월 1일부터 11월 31일
 - 기준 시점 : 매년 7월 만근한 근로자

(3) 조사현황

○ 2010년도 엔지니어링업체 임금실태조사에서는 2010년 6월말 기준 엔지니어링협회에 신고된 4,447개 사업체를 대상으로 조사를 실시한 결과, 소재불명(우편물반송)으로 판단되는 1,077개 사업체가 조사대상에서 제외되었다. 나머지 3,370개 사업체 중 2,817개 사업체에서 조사표가 회수되었으며 이 중 항목무응답 등 불성실한 조사표 109개를 제외하고 최종적으로 2,708개 사업체를 집계에 사용하였다.

<표 2> 회수 및 집계 현황

(단위 : 개)

구 분	집계대상 업 체 수	자료회수		자료집계	
		회수현황	회수율(%)	집계현황	집계율(%)
기 계	173	136	78.6	135	78.0
선 박	15	11	73.3	11	73.3
항 공·우 주	1	1	100.0	1	100.0
금 속	4	2	50.0	2	50.0
전 기·전 자	150	100	66.7	98	65.3
통신정보처리	654	470	71.9	439	67.1
화 학	27	23	85.2	22	81.5
광 업 자 원	3	1	33.3	1	33.3
건 설	1,980	1,782	90.0	1,717	86.7
환 경	117	92	78.6	86	73.5
농 립	114	92	80.7	91	79.8
해 양 수 산	32	29	90.6	29	90.6
산 업 관 리	12	6	50.0	6	50.0
응 용 이 학	88	72	81.8	70	79.5
소 계	3,370	2,817	83.6	2,708	80.4

※2010년도 엔지니어링업체 임금실태조사

2절 개선과제

(1) 현황

□ 모집단

목표모집단은 엔지니어링기술진흥법 제21조에 의해 신고된 업체이므로 모집단이 명확히 정의되어있다. 하지만 조사모집단은 목표모집단 가운데 소재불명 사업체가 조사대상에서 제외되므로 목표모집단과 조사모집단은 약간 달라질 수 있다.

□ 표본관리

전수조사이므로 표본관리 및 표본규모에 대한 해당사항은 없으며 표본오차와 표본추출방법에 대한 제시가 필요하지 않다.

(2) 문제점

조사대상 사업체 중 80%정도가 조사되고 있으며 조사되지 못하는 나머지 20%정도는 영세한 사업체로서 조사결과에 큰 영향을 주지 못하는 것으로 담당자들은 인식하고 있으나 조사되지 않은 사업체에 대하여 사후가중치 적용이나 보정등이 시행되지 않고 있다.

그리고 전수조사를 원칙으로 하다보니 조사예산을 고려하여 가능한 조사방법을 우편조사로 실시 하였다. 따라서 조사기간이 길어질 뿐만 아니라 조사대상 사업체에 대한 관리가 잘 되지 않는 문제점이 발생한다.

또한, 우편조사로 인한 자계식 기입방법으로 조사를 진행하다보니 조사 항목에 대한 이해가 부족한 경우 작성 오류를 범할 가능성을 가지고 있다.

(3) 개선사항

현재 이루어지고 있는 전수조사에 대하여 앞서 제시한 바와 같은 문제점들이 발생하므로 모집단에 대한 체계적인 분석이 선행되고 표본조사로의 전환여부를 판단하여 현실가능한 표본설계 방안을 제시하는 것이 필요하다.

3절 유사조사 분석

본 조사에 대해 2007년 정기품질진단 결과에서 제시된 개선사항 중 「측량업체임금실태조사」와 「건설업체임금실태조사」, 「감리원임금실태조사」 등의 조사들과 성격이 유사하여 정보교류의 필요성과 협조체제 구축을 강조하였다. 본 조사와 가장 유사하고 표본조사로 시행되고 있는 조사로 「소프트웨어기술자임금실태조사」를 들 수 있다. 따라서 한국 소프트웨어산업협회의 「소프트웨어기술자임금실태조사」 표본설계 방법을 중심으로 살펴본다.

□ 조사명칭 : 소프트웨어기술자임금실태조사

□ 조사목적

- 소프트웨어기술자임금실태조사는 소프트웨어 기술(S/W)의 실지급 임금을 조사하여 그 결과를 S/W사업 수행시 투입 기술자의 노임 단가로 적용할 수 있도록 해당업체 및 유관기관에 제공하고 S/W 기술인력의 동향을 파악하기 위한 기초자료로 활용하는 것을 목적으로 한다.

□ 조사개요

- 조사대상 : S/W사업자로 신고한 업체로 IT서비스, 패키지를 포함한다.

- 조사방법 : 우편조사(이메일이나 방문조사도 병행)

□ 표본설계

- 모집단 : 소프트웨어사업자로 신고한 사업체 또는 한국소프트웨어 산업협회 정회원사에서 근무한 소프트웨어 전담기술자
- 표본추출틀 : 소프트웨어산업협회 정회원사로 신고한 업체
- 표본규모 : 1,300개
- 표본추출
 - 협회에 신고된 6,000여개 신고업체 중 지역별, 자본금별, 기술자 수, 매출액, 사업 분야 등을 기준으로 표본규모 산정
 - 지역별 표본수와 층별 표본배정은 확률추출법을 사용

□ 본 연구에 반영사항

- 조사 목적이나 조사항목이 유사하며 표본추출방법으로는 층화확률 비례추출법을 사용하였다.
- 본 연구에서도 업종별 종사자 규모별로 층화하여 표본규모를 산정 하도록 한다.
- 소프트웨어기술자임금실태조사에서는 업종별 분포가 반영되지 않는 확률추출법을 사용하였으나 본 연구에서는 업종별 분포를 파악하여 추출방법을 결정한다.

제3장 표본설계

1절 모집단 분석

(1) 모집단 설정

□ 목표모집단

본 연구에서의 목표모집단은 2011년 6월 기준 엔지니어링기술진흥업 시행령 제21조에 따라 한국엔지니어링협회에 신고한 사업체로 총 4,702개 사업체이다. 엔지니어링협회에서는 엔지니어링활동을 하는 사업체를 대상으로 14개 업종으로 모집단을 구분하였다.

<표 3> 산업별 · 종사자규모별 모집단 현황

(단위 : 개)

구 분	사업체수	구성비(%)
기 계	254	5.4
선 박	19	0.4
항 공·우 주	2	0.0
금 속	5	0.1
전 기·전 자	212	4.5
통신정보처리	817	17.4
화 학	38	0.8
광 업 자 원	3	0.1
건 설	2821	60.0
환 경	160	3.4
농 립	188	4.0
해 양 수 산	46	1.0
산 업 관 리	21	0.4
응 용 이 학	116	2.5
합 계	4,702	100.0

□ 조사모집단

조사모집단은 한국엔지니어링협회에 신고된 사업체 중 소재가 불명확하거나 조사가 불가능한 사업체를 제외하고 조사가 가능한 사업체를 대상으로 조사모집단을 결정한다.

□ 표본추출틀(Sampling Frame) 검토

표본추출틀로는 한국엔지니어링협회에서 신고된 자료로서 소재불명 등의 이유로 조사모집단에 포함되지 않는 사업체를 제외한 3,625개 사업체 자료를 표본추출틀로 사용하며 엔지니어링활동 사업체의 종사자수, 매출액 등 산업특성을 나타낼 수 있는 항목을 포함하고 있다.

따라서 엔지니어링산업 표본설계에서 산업의 형태를 잘 나타낼 수 있는 자료를 활용하는 것이 적합한가를 검토한다.

검토결과, 표본추출틀로 사용가능한 사업체 중 매출액을 기재하지 않은 사업체가 3,881개 사업체이며 종사자수를 기재하지 않은 사업체는 297개 사업체로 나타났다.

□ 표본추출틀 선정

표본추출틀을 준비하기 위해서는 추출단위(sampling unit)를 먼저 결정해야 한다. 표본틀이란 추출단위의 목록이기 때문이다. 따라서 표본추출틀은 조사의 특성에 적합하면서도 모집단에 포함된 조사단위들의 중복이나 누락을 최소화할 수 있어야 한다. 또한 표본틀의 설정, 이용, 유지 및 보완 등 실제 적용이 가능하고 예산에 무리가 따르지 않는 범위 내에서 이루어져야 한다.

엔지니어링 산업에서 표본설계를 위한 표본추출틀을 선정하기 위해 먼저 조사모집단의 정의에 따라 소재불명인 사업체 1,077개사를 제외한다.

그리고 엔지니어링협회에 신고된 자료 중 종사자수를 기재하지 않은 사업체에 대해서는 통계청의 “전국사업체조사”에서 사업자등록번호, 사업체명, 대표자명으로 신고된 자료와 연계하여 동일한 사업체로 판단 되는 경우 종사자수를 보완하여 표본추출틀을 작성한다.

<표 4> 표본추출틀 검토사항

(단위 : 개)

구 분	종사자수 '미기재' 사업체	매출액 '미기재' 업체	소재불명
기 계	12	200	57
선 박	0	16	4
항 공·우 주	0	1	1
금 속	0	4	1
전 기·전 자	12	160	49
통신정보처리	26	611	155
화 학	3	31	10
광 업 자 원	0	1	
건 설	203	2407	713
환 경	4	119	30
농 립	29	182	20
해 양 수 산	2	35	10
산 업 관 리	3	16	5
응 용 이 학	3	98	22
합 계	297	3,881	1,077

□ 보조변수 결정

본 표본설계에 사용할 기준 변수를 선정하기 위해 매출액, 종사자수를 살펴본 결과, 매출액으로 기준 변수를 선정하여 산업특성을 파악하는 것이 가장 적절하지만 엔지니어링협회에 신고된 업체 중 매출액이 미기재된 사업체가 대부분이므로 기준변수로 사용하기에는 정확성이 떨어진다.

따라서 산업크기를 측정할 수 있고 자료의 신뢰성이 높은 종사자수를 기준변수로 선정하여 표본설계에 사용한다.

□ 조사모집단 현황

신고된 자료 중 종사자수를 미기재한 사업체에 대해 종사자수를 보완하여 최종 조사모집단을 설정한다. 표본설계시 실제로 사용할 조사모집단의 종사자수가 어떠한 분포를 가지는가를 살펴보기 위해 업종별, 종사자 규모별을 나누어 살펴본다.

<표 5> 조사모집단 현황

(단위 : 개)

구 분	전체	종사자 규모별 사업체수						
		1-4인	5-9인	10-19인	20-49인	50-99인	100-199인	200인이상
기 계	197	12	18	51	38	28	18	32
선 박	15	-	-	6	2	2	1	4
항 공·우 주	1	-	-	-	-	-	1	-
금 속	4	-	3	-	1	-	-	-
전 기·전 자	163	11	9	37	54	24	14	14
통신정보처리	662	27	81	213	192	62	35	52
화 학	28	3	2	3	5	4	3	8
광 업 자 원	3	-	-	3	-	-	-	-
건 설	2,108	182	529	678	436	133	54	96
환 경	130	4	11	47	47	14	3	4
농 립	168	33	110	17	6	-	-	2
해 양 수 산	36	2	17	10	6	1	-	-
산 업 관 리	16	3	2	2	3	2	-	4
응 용 이 학	94	3	10	30	30	6	4	11
합 계	3,625	280	792	1,097	820	276	133	227

(2) 층화

표본설계에서 조사목적에 부합하도록 효과적인 층화를 하는 것이 무엇보다도 중요한 일이다. 층화란 모집단을 특성에 따라 서로 동질적인 몇 개의 그룹(층)으로 나누는 과정이다. 모집단을 몇 개의 층으로 나누 후에 각 층별로 독립적인 표본추출이 이루어진다.

엔지니어링산업은 14개 업종별 추정이 가능하도록 층화한다. 따라서 업종별로 1차 층화를 하고 업종별 추정값의 정확성을 높이기 위해 업종별 종사자수를 고려한 2차 층화를 실시한다.

이때, 업종별 업체수가 20개 이하인 업종인 선박, 항공·우주, 금속, 광업자원, 산업관리 업종에 대해서는 전수층으로 구분하였다.

그리고 산업부문 중 산업공장과 원자력발전 활동을 하는 사업체가 많이 포함된 기계, 전기·전자, 화학, 응용이학 업종에 대해서도 전수층으로 구분하였다.

<표 6> 업종별 사업체수 및 전수/표본 구분

구분	기계	선박	항공·우주	금속	전기·전자	통신정보처리	화학	광업자원	건설	환경	농림	해양수산	산업관리	응용이학
사업체수	197	15	1	4	163	662	28	3	2108	130	168	36	16	94
전수/표본	전수	전수	전수	전수	전수	표본	전수	전수	표본	표본	표본	표본	전수	전수

□ 층화구조

○ 1차 층화 : 14개 업종별로 1차 층화

- 엔지니어링협회에서 신고된 자료로 기계, 선박, 항공·우주, 금속, 전기·전자, 통신정보처리 등의 14개 업종으로 구분한다.

○ 2차 층화 : 종사자 규모별로 2차 층화

- 종사자규모는 종사자수에 따른 사업체 분포를 고려하여 1~9인, 10~19인, 20~49인, 50인이상의 4개 규모로 구분한다.

<표 7> 업종별 · 종사자규모별 구분기준

업종	1규모	2규모	3규모	4규모
통신정보처리, 건설, 환경, 농림, 해양수산	1-9인	10-19인	20-49인	50인이상

(3) 조사모집단 현황

조사모집단을 층화하여 사업체 현황을 살펴본다.

<표 8> 조사모집단 현황

(단위 : 개)

구분	종사자규모				합계
	1규모	2규모	3규모	4규모	
기 계	30	51	38	78	197
선 박	-	6	2	7	15
항 공·우 주	-	-	-	1	1
금 속	3	-	1	-	4
전 기·전 자	20	37	54	52	163
통신정보처리	108	213	192	149	662
화 학	5	3	5	15	28
광 업 자 원	-	3	-	-	3
건 설	711	678	436	283	2,108
환 경	15	47	47	21	130
농 립	143	17	6	2	168
해 양 수 산	19	10	6	1	36
산 업 관 리	5	2	3	6	16
응 용 이 학	13	30	30	21	94
소 계	1,072	1,097	820	636	3,625

2절 표본설계

(1) 표본크기

□ 표본크기

업종별, 규모별 추정치들에 대한 표본오차 뿐만 아니라 분석단위에 대한 표본오차 관리를 위해 업종별, 종사자규모별 목표오차를 설정하여 표본크기를 결정하여야 한다. 업종별 사업체수 규모에 따라서 상대표본오차 개념인 변동계수를 사용하여 각 층의 목표오차를 설정한다.

또한, 각 층을 부차모집단으로 생각하고 부차모집단별 목표오차를 정해 표본규모를 구하고 이들 부차모집단의 표본수를 합쳐 전체 표본규모를 산정한다. 이때 업종별 업체수가 20개 이하인 업종과 산업공장, 원자력발전과 관련된 업종에 대해서는 전수층으로 구분하여 표본규모를 결정한다.

목표오차를 달리 하였을 경우의 표본수는 아래의 <표 9>와 같다.

- 업종별 사업체수가 20개 이하인 경우와 산업공장, 원자력발전과 관련된 업종은 전수조사한다.
- 표본 규모는 통계작성기관인 한국엔지니어링협회와의 협의를 통해 최종 분석단위와 예산, 목표오차를 고려하여 결정한다.

표본크기를 결정할 때는 조사목적과 여건에 맞도록 표본의 크기를 결정해야 한다. 표본의 크기는 조사의 정밀도와 신뢰도 뿐만 아니라 조사에 들어가는 시간과 인력에도 영향을 끼치므로 적절하게 결정되어야 한다.

□ 표본배분

업종분류별 목표오차를 충족시키기 위해서는 각 층을 독립적인 모집단으로 보고 각 층별로 요구되는 목표정도에 따라 층내 종사자수의 표준편차를 이용하여 Neyman의 최적배분법에 의거하여 표본의 크기를 결정한다. 업종·종사자 규모별 표본수는 다음의 공식에 따라 결정되었다.

$$n = \left(\frac{C}{C_y} \right)^2 / \left[1 + \frac{1}{N} \left(\frac{C}{C_y} \right)^2 \right]$$

여기서 $C = S/\bar{Y}$: 모집단 변동계수(종사자수)

C_y : 목표오차

N : 업종·종사자 규모별 모집단 업체수

<표 9> 목차오차에 따른 표본수

(단위 : 개)

구 분	전수/ 표본	모집단수	목표오차에 따른 표본크기					
			2.0%	3.0%	5.0%	7.0%	10.0%	
기 계	전수	197	197	197	197	197	197	197
선 박	전수	15	15	15	15	15	15	15
항 공·우 주	전수	1	1	1	1	1	1	1
금 속	전수	4	4	4	4	4	4	4
전 기·전 자	전수	163	163	163	163	163	163	163
통신정보처리	표본	662	405	327	251	214	181	
화 학	전수	28	28	28	28	28	28	28
광 업 자 원	전수	3	3	3	3	3	3	3
건 설	표본	2,108	904	645	407	296	205	
환 경	표본	130	105	88	64	51	40	
농 립	표본	168	141	118	78	53	32	
해 양 수 산	표본	36	33	32	27	21	15	
산 업 관 리	전수	16	16	16	16	16	16	16
응 용 이 학	전수	94	94	94	94	94	94	94
소 계		3,625	2,109	1,731	1,348	1,156	994	

□ 표본크기 결정

표본의 크기가 커질수록 조사의 정확도는 높아진다. 그러나 표본이 커지면 조사비용이나 조사를 위한 노력이 많이 들기 때문에 경우에 따라 오히려 조사의 질을 떨어뜨리는 요인이 될 수도 있다. 그러므로 표본의 크기는 조사목적에 맞는 목표정도(target precision)를 정한 후 그것을 만족시키는 범위 내에서 정하는 것이 바람직하다.

엔지니어링협회에서 공표하고 있는 최종 분석단위가 엔지니어링산업 전체임을 고려할 때 업종별 허용오차가 5.0%일 때의 표본크기가 가장 적절한 것으로 판단된다.

<표 10> 표본수(목표오차 5.0%)

(단위 : 개)

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	251	64	13	27	147
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	407	128	16	27	236
환 경	64	14	11	19	20
농 립	78	64	7	5	2
해 양 수 산	27	15	7	5	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	1,348	361	186	216	585

※부록: 목표오차별 표본크기

(2) 표본추출

□ 표본추출

표본추출틀에서 업종·규모별 층화를 한 후 각 층에 표본을 할당하고, 할당된 표본의 크기로 표본을 추출하기 위해 적절한 분류변수를 선정하여 분류변수에 따라 사업체를 정렬하여 계통추출법으로 표본을 추출한다. 이 때 분류변수로 종사자수를 고려한다.

계통추출을 위해서 각 층에 속한 사업체를 종사자수로 정렬하여 계통추출한다.

□ 예비표본추출

표본사업체가 유고(전업, 휴·폐업, 불응 등)시 표본사업체와 동일업종 및 규모가 유사한 사업체로 선정되도록 표본사업체의 앞, 뒤 사업체를 예비표본으로 선정하며 유고사업체에 대한 표본대체 비율은 20%가 넘지 않도록 권장한다. 과도한 표본대체로 인하여 과소 또는 과대추정이 될 수 있으며 엔지니어링 산업특성이 제대로 반영되지 않을 수 있다.

(3) 모의실험

□ 개요

한국엔지니어링협회의 2010년 임금실태조사 조사결과를 토대로 본 연구에서 제시한 표본설계방법에 따라 모의추출을 실시하여 추정된 임금단가와 2010년에 공표된 임금단가와 오차를 비교하였다.

□ 모의실험 결과

- 2010년 조사사업체 : 2,708개 사업체
- 모의추출사업체 : 1,098개 사업체

<모의 실험 결과>

구분		2010년 공표임금	추정값	오차
건설 및 기타	기술사	325,979	320,224	-1.8%
	특급기술자	258,726	255,883	-1.1%
	고급기술자	203,802	199,944	-1.9%
	중급기술자	174,250	174,323	0.0%
	초급기술자	131,853	135,358	2.7%
	고급기능사	136,699	135,476	-0.9%
	중급기능사	138,346	140,371	1.5%
	초급기능사	111,171	112,833	1.5%
산업공장	기술사	367,499	368,746	0.3%
	특급기술자	324,978	325,928	0.3%
	고급기술자	248,131	248,735	0.2%
	중급기술자	208,469	209,874	0.7%
	초급기술자	161,124	162,735	1.0%
	고급기능사	165,162	165,268	0.1%
	중급기능사	142,143	142,622	0.3%
	초급기능사	116,627	116,735	0.1%
원자력발전	기술사	402,579	401,215	-0.3%
	특급기술자	360,629	361,081	0.1%
	고급기술자	299,132	298,681	-0.2%
	중급기술자	249,103	249,971	0.3%
	초급기술자	189,727	192,244	1.3%
	고급기능사	292,388	292,388	0.0%
	중급기능사	209,391	208,979	-0.2%
	초급기능사	145,781	145,781	0.0%

3절 추정

(1) 가중값

각 표본단위들이 모집단 단위를 얼마나 대표하는지를 조정하기 위해 가중값을 계산한다. 엔지니어링임금실태조사의 경우는 모집단 정보가 계속 변하기 때문에 가중값을 통해 모수를 추정하는 것이 매우 중요하고 어려운 일이다.

통계조사에서 가중값을 적용하는 이유는 첫째, 표본설계특성에 따라 존재하는 불균등한 표본단위의 추출확률을 조정하고 둘째, 표본단위의 무응답을 보정하고 셋째, 모집단의 알려진 특성을 일치시키기 위해 보정하기 위함이다.

일반적으로 표본조사 자료를 분석할 때 활용되는 가중값은 크게 다음과 같이 3가지로 계산하여 이용된다.

① 기본 가중값 부여

- 표본설계 특성을 나타내는 가중값을 의미한다.
- 표본추출단위의 추출확률에 대한 역수로 가중값을 계산한다.
- 표본추출방법이 다단계추출인 경우 각 단계별 추출단위의 추출확률을 역수로 계산하여 최종 표본추출단위의 가중값을 계산한다.

② 무응답 보정

- 최초 표본단위가 단위무응답인 경우 무응답 표본단위를 보정하기 위한 가중값으로 최초목표 표본 크기를 유효 표본크기로 나누어 조정한다.
- 최초표본단위를 대체표본으로 교체(substitution)하는 경우 무응답 보정 단계는 필요하지 않다.
- 항목 무응답은 대체(imputation)로 추정하여 사용 가능하지만 단위

무응답은 가중값단계에서 보정해야 한다.

③ 사후 층화 가중값 부여

- 사후 층화 가중값은 기본가중값과 무응답보정으로 적용된 가중값으로 파악된 모집단의 특성과 실제 알려진 모집단의 특성을 비교하여 중요도가 높은 속성의 차이를 보정하기 위해 적용되는 가중값을 의미한다.
- 사후 층화 가중값은 모집단에 대해 추정된 결과를 통해 실제 알려진 (대개 전수조사의 결과) 분포와 비교하기 때문에 모집단의 알려져 있는 중요한 변수를 선정하여 적용하면 된다.

(2) 추정

□ 전체 모집단에 대한 추정

- 추정에 사용될 기호들은 다음과 같다.

N_h : h 층의 모집단 사업체 수,

n_h : h 층의 표본 사업체 수,

$w_{hi} = \frac{N_h}{n_h}$: h 층의 i 번째 표본 사업체의 가중치

y_{hi} : h 층의 i 번째 표본 사업체로부터 얻은 측정값

- 전체 모집단에 대한 평균임금의 추정치인 평균(\bar{y})은 다음과 같다.

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

- 평균 임금에 대한 추정치인 \bar{y} 의 분산추정치는 다음과 같다.

$$var(\bar{y}) = \sum_{h=1}^L \frac{n_h(1-f_n)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_h)^2$$

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

- 전체 모집단에 대한 평균 임금의 추정치인 \bar{y} 의 상대표준오차(CV)는 다음과 같다.

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

□ 부차모집단(층)에 대한 추정

- h 번째 부차모집단(층)에 대한 평균 임금의 추정치 \bar{y}_h 와 이에 대한 분산추정치는 각각 다음과 같다.

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} w_{hi} y_{hi}}{\sum_{i=1}^{n_h} w_{hi}}, \quad var(\bar{y}_h) = \frac{n_h(1-f_n)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_h)^2$$

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

- 또한, \bar{y}_h 에 대한 상대표준오차는 다음과 같이 계산한다.

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

제4장 결론 및 제안

본 연구는 한국엔지니어링협회에서 매년 조사하는 엔지니어링업체 임금실태조사의 표본설계 개선을 위한 연구이다. 지난 2007년 실시한 통계청의 정기품질진단 결과에서 '표본설계' 부분에 대해 제기된 문제점에 대해 개선안을 제공하는데 그 목적이 있다.

□ 결론

- 2010년 엔지니어링업체 임금실태조사에 대한 특성을 살펴보고 표본조사로 전환이 가능한지를 검토한다.
- 검토결과, 엔지니어링협회에 신고된 자료를 검토하여 조사방법을 표본조사 전환할 수 있도록 방법론을 제시한다.
 - 먼저 신고된 자료를 토대로 모집단을 정의하였고 표본설계에 활용가능한 표본추출틀을 마련한다.
 - 업종별 추정이 가능하도록 14개 업종별로 1차 층화를 실시하고 업종별 추정값의 정확성을 높이기 위해 종사자 규모에 따라 2차 층화한다.
 - 업종 중 모집단 규모가 작은 업종은 전수조사를 그대로 유지하도록 하고 나머지 업종에 대해서는 업종별 표본규모를 제시한다.
 - 각 층의 표본이 적절하게 추출되도록 종사자수를 정렬하여 추출하는 계통추출방법을 제시한다.
 - 추출된 표본에 대해 조사완료 후 추정이 가능하도록 가중치 산정방법을 제시하고 모집단에 대한 추정식을 제시한다.
- 본 연구를 통해 조사방법을 전수조사에서 표본조사로의 전환에 따른 정확성과 신뢰성에 대해 불확실성이 사라질 것으로 예상된다.

□ 제안

- 본 연구를 수행하면서 모집단 및 표본추출틀을 분석한 후, 다음 몇 가지 사항들에 대한 추가적인 보완이 이루어진다면 전반적으로 통계의 품질이 더욱 향상 시킬 것으로 기대한다.
 - 정확한 표본조사가 이루어지기 위해서는 표본추출틀에 대한 자료의 정확성이 요구된다. 매년 신고되는 엔지니어링활동 사업체에 대한 종사자, 매출액 등의 자료를 정확하게 기재되어야한다.
 - 산업 규모를 결정하기 위해서는 종사자수보다는 매출액이 산업의 크기를 잘 나타낼 수 있는 변수이므로 엔지니어링협회에서 신고된 사업체에 대해 조사 또는 보고 등으로 매출액에 대한 자료를 구축할 필요성이 있다.
 - 향후 엔지니어링산업의 특성이 변화함에 따라 주기적으로 산업의 특성이 잘 반영된 새로운 표본설계를 해야 할 것이다.

[참고문헌]

- 2010년도 엔지니어링업체 임금실태조사(한국엔지니어링협회, 2010)
- 표본조사의 이해와 활용(김영원·류제복·박진우·홍기학, 2009)
- 축산물생산비조사를 위한 복합표본설계(김수택, 2008)
- 표본조사를 위한 설계과정(류제복, 2008)
- 노동통계조사를 위한 표본설계(매월노동통계조사, 노동력수요동향조사를 중심으로-, 응용통계연구 제10권 2호, p215-226.)
- Sampling Techniques, Wiley(Cochran, W. G., 1977)
- Survey Sampling, Wiley(Kish, L., 1965)
- Sampling: Design and Analysis, Duxbury Press(Lohr, S., 1999)
- SAS/STAT User's Guide, Version 8(SAS, 1999)

[부록 : 목표오차별 표본크기]

<목표오차 2% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	405	97	63	96	149
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	904	411	91	128	274
환 경	105	15	31	38	21
농 립	141	119	14	6	2
해 양 수 산	33	18	9	6	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	2,109	736	340	407	626

<목표오차 3% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	327	87	33	59	148
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	645	269	44	68	264
환 경	88	15	22	30	21
농 립	118	99	11	6	2
해 양 수 산	32	17	9	6	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	1,731	563	251	302	615

<목표오차 4% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	282	75	20	39	148
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	498	181	25	41	251
환 경	74	14	15	24	21
농 립	96	80	9	5	2
해 양 수 산	29	16	8	5	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	1,500	442	209	247	602

<목표오차 5% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	251	64	13	27	147
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	407	128	16	27	236
환 경	64	14	11	19	20
농 립	78	64	7	5	2
해 양 수 산	27	15	7	5	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	1,348	361	186	216	585

<목표오차 7% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	214	46	7	15	146
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	296	71	8	14	203
환 경	51	13	6	12	20
농 립	53	42	5	4	2
해 양 수 산	21	12	5	4	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	1,156	260	163	182	551

<목표오차 10% 표본크기>

구 분	합계	종사자규모			
		1규모	2규모	3규모	4규모
기 계	197	30	51	38	78
선 박	15	0	6	2	7
항 공·우 주	1	0	0	0	1
금 속	4	3	0	1	0
전 기·전 자	163	20	37	54	52
통신정보처리	181	29	3	7	142
화 학	28	5	3	5	15
광 업 자 원	3	0	3	0	0
건 설	205	37	4	7	157
환 경	40	11	3	7	19
농 립	32	24	3	3	2
해 양 수 산	15	9	3	3	0
산 업 관 리	16	5	2	3	6
응 용 이 학	94	13	30	30	21
소 계	994	186	148	160	500