

발 간 등 록 번 호

11-1240000-000695-01

2013년도 국가통계 품질개선 컨설팅 연구용역

『연구개발활동조사』  
품질개선 컨설팅 최종결과보고서  
- 통계 작성 개선 연구-

2013. 12.

제 출 문

# 제 출 문

통계청장 귀하

본 보고서를 「연구개발활동조사」 품질개선 컨설팅의 최종결과보고서로 제출합니다.

2013년 12월

한국통계학회 ㉞

## 연구진

---

책임연구원	이기재, 한국방송통신대학교	정보통계학과	교수
연구보조원	박원서, 한국방송통신대학교	정보통계학과	조교
	김윤기, 한국방송통신대학교	정보통계학과	조교
통 계 청	김 원 태, 통계청 품질관리과		사무관
	윤 은 경, 통계청 품질관리과		주무관

요약문

## 최종결과보고서 요약문

연구과제명	연구개발활동조사 통계 작성 개선 연구
주제어	연구개발활동조사, 표본조사, 무응답, 표본추가
연구기간	2013. 10. 18. - 2013. 12. 09.
연구기관	한국통계학회
연구진구성	이기재, 한국방송통신대학교 정보통계학과 교수
<p>「연구개발활동조사」는 우리나라의 과학기술 연구 개발 활동(연구개발인력 및 연구개발비 등) 현황을 조사하여 국가 과학기술 정책 수립 등에 필요한 기초자료를 제공하고, 각계의 과학기술 분야 종사자로 하여금 과학기술 연구개발 계획 등에 참고자료로 제공하는 것을 목적으로 한다.</p> <p>본 연구는 현행 「연구개발활동조사」의 표본조사 전환의 타당성과 효과를 살펴보고, 조사 개선 방안을 제시함으로써 추정치 통계적 정확도를 높이는 것을 목적으로 한다. 최근 3년 동안의 현행 조사 데이터를 분석하여 표본설계와 조사 개선 방안을 제시하였다. 본 연구의 주요 결과를 요약하면 다음과 같다.</p> <p>첫째, 조사 회수율 현황은 전반적으로 약간씩 떨어지는 추세로 나타났다. 공공연구기관, 대학, 의료기관 등의 회수율은 전반적으로 97%를 상회하고 있지만, 기업체의 회수율은 80% 초반의 값을 보이다가 2011년도 조사에서는 77.9%를 나타내고 있다. 기업체의 회수율은 공공연구기관, 대학, 의료기관과 비교하여 상당히 낮은 수준을 보이고 있다.</p> <p>둘째, 「연구개발활동조사」에서 기업체의 연구개발비와 연구개발인력 측면에서 차지하는 비중은 각각 약 75%과 약 55%를 보이고 있다. 기업체 규모에 따라 연구개발비와 연구개발인력에 큰 차이가 있어, 두 조사항목에 대해서 오른쪽으로 꼬리가 긴 비대칭적 분포를 갖는다.</p> <p>셋째, 모집단에 속한 기업체는 36,404개소이고, 기업체 규모별 현황을 보면 소기업이 18,951개소(51.1%)과 벤처기업이 9,885개소(27.2%)이고, ‘제조업’에 25,235개소(69.3%), ‘출판, 영상, 방송통신 및 정보서비스업’에 4,741개소(13.0%), ‘건설업’에 2,502개소(6.9%), ‘전문 과학 및 기술서비스업’에 2,110개소(5.8%) 등으로 집중되어 있다. ‘제조업’과 ‘출판, 영상, 방송통신 및 정보서비스업’의 벤처기업 또는 소규모 기업체가 24,248개소로 전체 모집단 기업체의 66.6%를 차지하고 있다.</p> <p>넷째, 「연구개발활동조사」에 대한 조사 개선 방안으로 표본추출틀의 정비, 특이치 검출 및 처리 방법, 조사데이터 신뢰성 검토를 위한 사후 표본조사 점검과 무응답 기업체에 대한 부차 표본조사 등을 제시하였다.</p> <p>다섯째, 「연구개발활동조사」에 대한 표본조사 대상을 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 100인 미만 소규모 기업체로 국한할 것을 제안하였고, 표본추출방안에 따라 표본조사의 타당성과 효과를 살펴보았다.</p>	

## 차 례

제 1 장 서 론 .....	1
1. 연구배경 .....	1
2. 연구내용 및 방법 .....	2
제 2 장 현행 조사데이터 분석 .....	4
1. 현행 조사의 응답 현황 분석 .....	4
2. 현행 조사의 기관유형별 현황 분석 .....	6
3. 연구개발비 및 연구개발인력 현황 분석 .....	10
4. 기업체의 연구개발비 및 연구개발인력 현황 분석 .....	15
5. 조사 개선을 위한 시사점 .....	19
제 3 장 모집단 분석 .....	20
제 4 장 현행 조사 개선 방안 .....	25
1. 모집단 구성 .....	25
2. 특이치 검출 및 처리 방안 .....	26
3. 조사데이터의 신뢰성 검토를 위한 조사 방안 .....	33
제 5 장 표본조사 적용 방안 .....	36
1. 표본조사 방식을 적용할 때 고려사항 .....	36
2. 모집단 분석: 표본조사 대상 .....	37
3. 표본크기 결정 .....	39
4. 표본추출 방안 .....	42
5. 추정 방법 .....	43

제 6 장 결론 및 요약 .....	45
1. 연구결과 요약 .....	45
2. 통계 작성 개선 로드맵 .....	46
참고문헌 .....	48
부록 1. 2011년도 「연구개발활동조사」의 산업중분류 및 규모별 응답 기업체 현황 .....	51
부록 2. 2011년도 「연구개발활동조사」의 산업중분류 및 규모별 분석 결과 .....	53
부록 3. 산업중분류 및 기업체 규모별 모집단 현황 .....	57

## < 표 목 차 >

<표 1-1> 주요 조사내용 .....	1
<표 1-2> 조사대상 기관 현황: 2011년도 조사 .....	2
<표 2-1> 조사대상 기관 현황 .....	4
<표 2-2> 연구주체 유형 및 종업원 수 규모별 현황 .....	6
<표 2-3> 전체 응답기관 중 연구개발비가 양수인 기관 현황 .....	7
<표 2-4> 연도 및 규모별 기업체 현황 .....	8
<표 2-5> 연도 및 산업대분류별 기업체 현황 .....	8
<표 2-6> 연도 및 규모별 기업체 현황 .....	9
<표 2-7> 산업대분류 및 규모별 기업체 현황: 2011년도 .....	10
<표 2-8> 연도 및 연구주체별 연구개발비 현황 (단위: 억원) .....	10
<표 2-9> 연도 및 산업대분류별 연구개발비 현황 (단위: 억원) .....	10
<표 2-10> 연도 및 연구주체별 연구개발인력 현황 (단위: 인) .....	12
<표 2-11> 연도 및 기업체 규모별 연구개발인력 현황 .....	12
<표 2-12> 연도 및 산업대분류별 연구개발인력 현황 (단위: 인) .....	13
<표 2-13> 연구주체별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도 .....	13
<표 2-14> 기업체 규모별 조사대상 기관, 연구개발비, 연구개발인력 현황: 2011년도 .....	14
<표 2-15> 산업대분류별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도 .....	14
<표 2-16> 연도별 연구개발비 분포 현황: 전체 연구기관 (단위: 억원) .....	16
<표 2-17> 연도별 연구개발비 분포 현황: 기업체 대상 (단위: 억원) .....	16
<표 2-18> 연도별 연구개발인력 분포 현황 (단위: 인) .....	17
<표 2-19> 연도별 연구개발인력 분포 현황: 기업체 대상 (단위: 인) .....	17
<표 2-20> 기업체 규모별 평균 연구개발비 현황 (단위: 억원) .....	18
<표 2-21> 기업체 규모별 평균 연구개발인력 현황 (단위: 인) .....	18
<표 2-22> 기업체 규모별 연구개발비와 연구개발인력의 상관계수 .....	19
<표 3-1> 산업대분류 및 기업체 규모별 현황 .....	21
<표 3-2> 지역 및 기업체 규모별 현황 .....	22
<표 3-3> 업종 및 기업체 규모별 현황 .....	22
<표 3-4> 연구분야 코드 및 기업체 규모별 현황 .....	23
<표 3-5> 기업체 규모 및 연구소 인정 연도별 현황 .....	24
<표 5-1> 산업대분류별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도 .....	37
<표 5-2> 산업중분류 및 기업체 규모별 조사대상 기업체 현황: 2011년도 조사 .....	38
<표 5-3> 목표 상대표준오차 설정 방안 .....	39
<표 5-4> 각 산업중분류 및 기업체 규모별 표본크기 현황: 방안 4 .....	40
<표 5-5> 산업중분류 및 기업체 규모별 모집단 기업체 현황 .....	41
<표 6-1> 「연구개발활동조사」 통계 작성 개선 로드맵 .....	47

## <그림 목 차>

<그림 2-1> 각 연구주체 유형별 회수율 현황 .....	5
----------------------------------	---

# 제 1 장 서 론

## 1. 연구배경

「연구개발활동조사」는 우리나라의 과학기술 연구 개발 활동(연구개발인력 및 연구개발비 등) 현황을 조사하여 국가 과학기술 정책 수립 등에 필요한 기초자료를 제공하고, 각계의 과학기술 분야 종사자로 하여금 과학기술 연구개발 계획 등에 참고자료로 제공하는 것을 목적으로 한다. 또한 OECD에 우리나라 연구개발 활동 현황을 제공하여 국가 신뢰도 제고 및 국가간 비교자료로 활용되고 있다.

이 조사는 통계법에 의한 지정통계(승인번호 제10501호, 1982. 7. 16)로 1963년 「연구기관실태조사」라는 명칭으로 시작되어 매년 전년도 연구개발활동 실적을 조사하여 발표하고 있다. 「연구개발활동조사」의 주요 조사항목은 <표 1-1>과 같다. 주요 조사내용은 일반 현황, 연구인력 현황(성, 학위, 전공 등), 연구개발비 현황(연구개발 단계, 재원, 비목 등) 등으로 구성되어 있다. 조사 대상 시점은 인원, 자본금 등에 대해서는 전년도 12월 31일을 기준으로 하고, 매출액, 연구개발비 등은 전년도 1월 1일 ~ 12월 31일을 기준으로 하고 있다.

<표 1-1> 주요 조사내용

구분	주요 조사 항목
일반 현황	- 조직 형태, 총 직원 수 등 기관의 일반현황 - 연구개발활동 수행 여부
인력 현황	- 연구개발 관계 종사자 (연구원, 연구보조원, 실질 연구참여 인력) 현황 - 연구원의 학위별·전공별·성별·연령별 현황
연구개발비 현황	- 재원별·비목별·연구개발 단계별·기술분류별 사용연구개발비 현황 - 경제사회 목적별·6T별·연구 분야별 사용 연구개발비 현황
기 타	- 지역별 연구개발비 및 연구개발인력 현황 - 기업체의 연구기자재 보유 및 구입 실적 - 기업체의 기술수출 실적 - 기업체의 해외 과학기술인력 유치 실적 - 기업체의 주요 연구개발과제 수행 현황

「연구개발활동조사」는 국가승인통계(10501호)로서 통계법에 따라 비밀을 보장함을 밝히고 자료를 수집하고 있다. 조사방법은 우편 및 온라인을 통한 자계식(自計式) 조사를 기본으로 하고, 전화조사 등을 부가적으로 실시하고 있다.

이 조사의 조사대상은 전국의 공공연구기관, 대학, 의료기관, 기업체 등이다. <표 1-2>는 2012년도 조사에서 조사대상 기관 현황이다. 전체 조사대상 기관은 32,673개소인데, 이 중 기업체는 30,913개소로 전체 조사대상 기관의 94.6%로 거의 대부분을 차지하고 있다. 한편, 조사표 회수율 현황을 보면 전체 조사대상 기관 중 25,816개소가 응답하여 회수율은 79.0%로 나타났다. 다른 기관에 비해서 기업체의 회수율이 77.9%로 다소 떨어지고 있다.

<표 1-2> 조사대상 기관 현황: 2011년도 조사

구분	공공연구기관	대학	의료기관	기업체	전체
조사 대상 기관수	754	428	578	30,913	32,673
조사표 회수 기관 수	735	419	571	24,091	25,816
회수율	97.5%	97.9%	98.8%	77.9%	79.0%

이 조사는 앞서 살펴본 바와 같이 전수조사로 진행되고 있으나, 그 동안 표본조사 방식으로의 전환 필요성이 지속적으로 제기되어 왔고, 기관유형에 따라 조사표 회수율(응답률)에 상당한 편차를 보이고 있다.

본 연구는 현행 「연구개발활동조사」의 분석하여 전반적인 조사 현황과 주요 특징을 분석하고, 모집단 자료 분석을 통해서 모집단 대표성이 확보될 수 있는 표본설계 방안을 연구함으로써 성공적인 표본조사로 정착시키는 방안을 모색하는 데 있다.

## 2. 연구내용 및 방법

본 연구에서 다루어질 연구내용과 방법은 다음에 제시하는 다섯 가지 사항을 중심으로 「연구개발활동조사」의 표본조사 방식 전환에 대한 타당성과 효과를 검토한다. 이를 토대로 모집단 대표성과 추정결과의 통계적 정확도가 높은 표본설계 방안을 제안하고자 한

다.

□ 현행 「연구개발활동조사」의 무응답 현황 및 처리 방법, 추정방법 검토

현행 조사의 무응답 현황을 분석하여 무응답으로 인한 모집단 대표성의 훼손 정도를 파악하고, 현행 조사의 무응답 처리 및 추정법 등에 대해서 검토하는 것이 필요하다.

□ 현행 조사의 모집단 분석

「연구개발활동조사」에 대한 최신 모집단 자료로 「기업부설연구소 및 연구개발 전담부서 현황(2012년 12월 말 기준)」 결과를 분석함으로써 현행 조사의 모집단 대표성을 검토한다. 이 조사에 대한 모집단 자료를 분석함으로써 현행 조사의 모집단 대표성을 검토하고, 새로운 표본설계의 기본방향을 마련한다.

□ 주요 조사항목에 대한 통계 분석

현행 조사의 주요 조사항목에 대한 추정 결과와 현황을 살펴봄으로써, 궁극적인 목표인 새로운 표본설계의 표본 크기 및 층화, 표본배분법 등에 대한 기본원칙과 방향을 설정한다.

□ 표본설계 방안 마련

「연구개발활동조사」를 표본조사 방식으로 전환하기 위하여 층화, 표본배분, 표본크기, 표본추출법 등에 대해서 연구한다. 특히 이 조사의 주요 관심변수인 연구개발비의 분포 특성을 반영할 수 있는 Lavalée & Hidiroglou (1988) 방법에 기초한 변형절사추출법의 적용을 검토한다.

□ 가중치 및 추정방법 연구

앞선 표본조사 방식으로 전환됨에 따라 가중치 작성 및 추정방법에 대한 연구와 더불어 항목무응답에 대한 대체 방안에 대한 연구도 필요하다.

## 제 2 장 현행 조사데이터 분석

### 1. 현행 조사의 응답 현황 분석

2011년도 「연구개발활동조사」의 조사 대상 분야는 OECD의 「연구개발활동조사시행지침(FRASCATI Manual)」에 따른 이학, 공학, 의학, 농학, 인문학, 사회과학 분야이다. 2008년 조사(2007년 결과)부터 인문·사회과학분야 포함되었다. 조사 방법은 우편 및 온라인을 통한 자계식(自計式) 조사를 기본으로 하고, 전화조사 등을 부가적으로 실시하고 있다.

이 조사의 대상은 전국의 공공연구기관, 대학, 의료기관, 기업체 등이다. <표 2-1>은 최근 4개년도(2008년, 2009년, 2010년, 2011년) 조사의 조사대상 기관 및 회수율 현황이다. 전반적으로 조사대상 기관 및 회수된 기관은 매년 증가하고 있다. 특히 조사대상 기관 중에서 기업체 수는 매년 증가폭이 가장 커서 증가분의 대부분을 차지하고 있다. 반면 공공연구기관, 대학, 병원 등은 다소 감소하거나 변화가 없는 것으로 나타나고 있다. 전체 조사대상 기관과 응답 기관에서 기업체가 차지하는 비중은 90% 이상으로 절대적으로 높은 상황이다.

<표 2-1> 조사대상 기관 현황

① 2008년도 조사(조사대상 기간: 2008년 1월 1일 ~ 12월 31일)

구분	공공연구기관	대학	의료기관	기업체	전체
조사 대상 기관(A)	743	410	651	21,506	23,310
	3.2%	1.8%	2.8%	92.3%	100.0%
조사표 회수 기관(B)	733	409	649	17,328	19,119
	3.8%	2.1%	3.4%	90.6%	100.0%
회수율(B/A)	98.7%	99.8%	99.7%	80.6%	82.0%

② 2009년도 조사(조사대상 기간: 2009년 1월 1일 ~ 12월 31일)

구분	공공연구기관	대학	의료기관	기업체	전체
조사 대상 기관(A)	752	431	635	23,874	25,692
	2.9%	1.7%	2.5%	92.9%	100.0%
조사표 회수 기관(B)	729	429	631	19,762	21,551
	3.4%	2.0%	2.9%	91.7%	100.0%
회수율(B/A)	96.9%	99.5%	99.4%	82.8%	83.9%

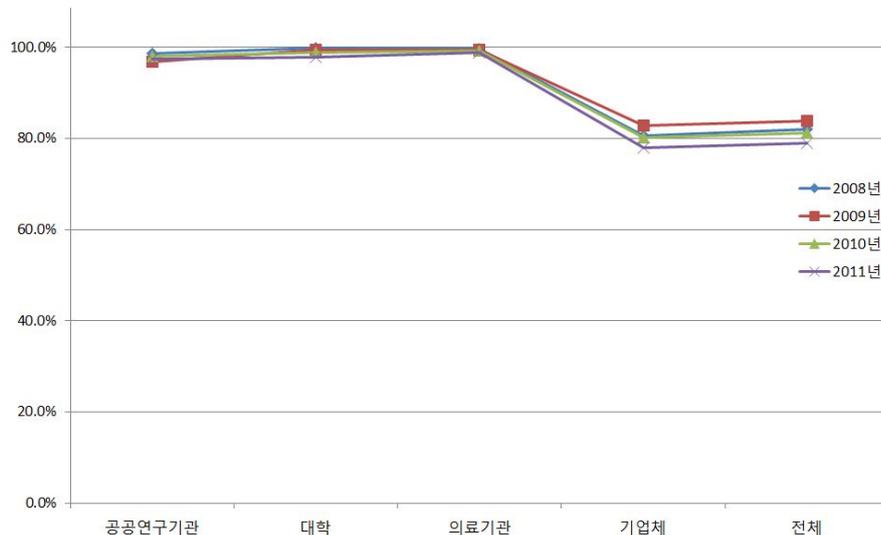
③ 2010년도 조사(조사대상 기간: 2010년 1월 1일 ~ 12월 31일)

구분	공공연구기관	대학	의료기관	기업체	전체
조사 대상 기관(A)	755	428	553	27,790	29,526
	2.6%	1.4%	1.9%	94.1%	100.0%
조사표 회수 기관(B)	740	423	549	22,250	23,962
	3.1%	1.8%	2.3%	92.9%	100.0%
회수율(B/A)	98.0%	98.8%	99.3%	80.1%	81.2%

④ 2011년도 조사(조사대상 기간: 2011년 1월 1일 ~ 12월 31일)

구분	공공연구기관	대학	의료기관	기업체	전체
조사 대상 기관(A)	754	428	578	30,913	32,673
	2.3%	1.3%	1.8%	94.6%	100.0%
조사표 회수 기관(B)	735	419	571	24,091	25,816
	2.8%	1.6%	2.2%	93.3%	100.0%
회수율(B/A)	97.5%	97.9%	98.8%	77.9%	79.0%

조사 회수율 현황을 보면 전반적으로 약간씩 떨어지는 추세인 것으로 보인다. 기업체의 회수율은 공공연구기관, 대학, 의료기관과 비교하여 상당히 낮은 수준을 보이고 있다. 연도별 변동은 거의 없고, 공공연구기관, 대학, 의료기관 등의 회수율은 전반적으로 97%를 상회하고 있지만, 기업체의 회수율은 80% 초반의 값을 보이다가 2011년도 조사에서는 77.9%를 나타내고 있다.



<그림 2-1> 각 기관유형별 회수율 현황

## 2. 현행 조사의 기관유형별 현황 분석

「연구개발활동조사」에서 공공연구기관은 국·공립기관, 정부출연기관, 지방자치단체출연기관, 기타비영리, 국·공립병원, 사립병원 등으로 구분된다. 여기서는 2009년-2011년도 조사에 응답한 기관의 연도별 현황을 분석한다. <표 2-2>는 연구개발비가 0보다 큰 응답기관의 현황을 분석한 것이다. 연구개발비가 있다고 응답한 전체 기관 중에서 100인 미만(종업원 수 기준) 규모의 기관 비율은 연도별 편차는 있지만 약 80%를 상회하고 있다.

<표 2-2> 연구주체 유형 및 종업원 수 규모별 현황

### ① 2009년도

구분	9인 이하	10-49인	50-99	100-299	300-999	1,000 +	결측	전체
공공연구기관	21	60	46	57	41	13	4	242
대학	5	25	41	115	87	34	8	315
의료기관	1	2	0	9	16	30	3	61
기업체	1,973	8,042	2,127	1,719	545	213	0	14,619
총합	2,000	8,129	2,214	1,900	689	290	15	15,237

### ② 2010년도

구분	9인 이하	10-49인	50-99	100-299	300-999	1,000+	결측	전체
공공연구기관	20	62	52	55	40	11	1	241
대학	3	25	37	116	89	37	6	313
의료기관	1	1	1	7	15	34	2	61
기업체	2,123	8,991	2,287	1,787	526	221	0	15,935
총합	2,147	9,079	2,377	1,965	670	303	9	16,550

### ③ 2011년도

구분	9인 이하	10-49인	50-99	100-299	300-999	1,000+	결측	전체
공공연구기관	13	59	50	59	47	14	4	246
대학	4	31	34	108	91	41	3	312
의료기관	0	0	2	11	15	35	1	64
기업체	2,957	11,323	2,712	1,980	553	225	0	19,750
총합	2,974	11,413	2,798	2,158	706	315	8	20,372

<표 2-3>은 최근 3개년도 조사를 대상으로 전체 응답한 기관 중 연구개발비가 0보다 크다고 응답한 기관 현황을 정리한 것이다. 공공연구기관, 대학, 의료기관 등에서 연구개발비가 있다고 응답한 사업체의 비율은 연도별 편차가 적은 것과 비교하여 기업체의 연구개발비가 있다고 응답한 비율은 상대적으로 편차가 큰 것으로 나타나고 있다. 특히 2011년도 조사에서 조사표가 회수된 기업체 중 연구개발비가 있다고 응답한 기업체의 비율은 82.0%로 이전 조사와 비교하여 상당히 높은 값을 보이고 있다. 이에 따라 2011년도 조사에서 전체 응답 기관 중 연구개발비가 있다고 응답한 기관의 비율은 78.9%로 2010년도 조사에 비해서 9.8%p 높았다. 기업체의 연구개발 활동이 경기 동향과 변동에 따라 연도별로 크게 차이날 수 있음을 시사하고 있다.

<표 2-3> 전체 응답기관 중 연구개발비가 양수인 기관 현황

구분	2009년도			2010년도			2011년도		
	조사표 회수(A)	연구개발비 유 기관(B)	비율 (B/A)	조사표 회수(A)	연구개발비 유 기관(B)	비율 (B/A)	조사표 회수(A)	연구개발비 유 기관(B)	비율 (B/A)
공공연구 기관	729	242	33.2%	740	241	32.6%	735	246	33.5%
대학	429	315	73.4%	423	313	74.0%	419	312	74.5%
의료기관	631	61	9.7%	549	61	11.1%	571	64	11.2%
기업체	19,762	14,619	74.0%	22,250	15,935	71.6%	24,091	19,750	82.0%
전체	21,551	15,237	70.7%	23,962	16,550	69.1%	25,816	20,372	78.9%

앞서 살펴본 바와 같이 전체 응답 기관 중에서 기업체가 차지하는 비중은 80% 이상으로 기업체가 차지하는 비중과 중요도는 대단히 높다. 따라서 기업체에 대한 세부 응답 현황을 분석하는 것은 향후 조사 개선방안을 찾는 데 있어서 대단히 중요할 것이다.

<표 2-4>는 최근 3개년 조사에 대해서 기업체의 규모별 응답 현황을 정리한 것이다. 전체 응답 기업체 수는 매년 증가하고 있고, 종사자 100인 미만의 규모에서 증가폭이 큰 것으로 나타나고 있다. 반면 종사자 300인 이상의 대규모에서 응답 기업체 수는 연도별 편차가 거의 없다. 기업체 대상의 「연구개발활동조사」는 일반적인 기업체 대상의 조사와 마찬가지로 기업체 수의 측면에서 보면 100인 미만의 기업체가 차지하는 비중이 매년 85% 내외를 보이고 있다.

<표 2-4> 연도 및 규모별 기업체 현황

구분	2009년도		2010년도		2011년도	
	기관 수	열%	기관 수	열%	기관 수	열%
10인 미만	1,973	13.5%	2,123	13.3%	2,957	15.0%
10-49인	8,042	55.0%	8,991	56.4%	11,323	57.3%
50-99인	2,127	14.5%	2,287	14.4%	2,712	13.7%
100-299인	1,719	11.8%	1,787	11.2%	1,980	10.0%
300-999인	545	3.7%	526	3.3%	553	2.8%
1000인+	213	1.5%	221	1.4%	225	1.1%
총합	14,619	100.0%	15,935	100.0%	19,750	100.0%

Note: 기관 수는 조사표 회수 기업체 중 해당 연도에 연구개발비가 있는 기업체 수임.

<표 2-5>는 최근 3개년 조사에 대해서 기업체의 산업대분류별 응답 현황을 정리한 것이다. 전체 응답 기업체 중에서 ‘제조업’에 속한 기업체의 비중이 가장 높게 나타나서 전체적으로 65%를 상회하고 있고, ‘출판, 영상, 방송통신 및 정보서비스업’의 비중이 11% 내외를 나타내고 있다. 다음으로 ‘건설업’과 ‘전문 과학 및 기술서비스업’의 비중이 높다. 이들 4개 업종의 비중은 전체 기업체의 약 94%이다.

<표 2-5> 연도 및 산업대분류별 기업체 현황

산업대분류	연도		2009년도		2010년도		2011년도	
	기관 수	열%	기관 수	열%	기관 수	열%	기관 수	열%
A. 농업, 임업 및 어업	0	0.0%	21	0.1%	34	0.2%		
B. 광업	0	0.0%	5	0.0%	0	0.0%		
C. 제조업	9,738	66.6%	10,488	65.8%	13,353	67.6%		
D. 전기, 가스, 증기 및 수도사업	11	0.1%	18	0.1%	35	0.2%		
E. 하수폐기물처리, 원료재생 및 환경복원	69	0.5%	71	0.4%	111	0.6%		
F. 건설업	1,322	9.0%	1,425	8.9%	1,626	8.2%		
G. 도매 및 소매업	627	4.3%	729	4.6%	755	3.8%		
H. 운수업	12	0.1%	13	0.1%	17	0.1%		
I. 숙박 및 음식점업	5	0.0%	5	0.0%	7	0.0%		
J. 출판, 영상, 방송통신 및 정보서비스	1,703	11.6%	1,893	11.9%	2,258	11.4%		
K. 금융 및 보험업	2	0.0%	4	0.0%	2	0.0%		
L. 부동산업 및 임대업	17	0.1%	10	0.1%	11	0.1%		
M. 전문, 과학 및 기술 서비스업	1,022	7.0%	1,143	7.2%	1,415	7.2%		
N. 사업시설관리 및 사업지원 서비스업	42	0.3%	42	0.3%	49	0.2%		
P. 교육 서비스업	22	0.2%	30	0.2%	37	0.2%		
Q. 보건업 및 사회복지 서비스업	1	0.0%	1	0.0%	1	0.0%		
R. 예술, 스포츠 및 여가관련 서비스업	3	0.0%	6	0.0%	9	0.0%		
S. 협회 및 단체, 수리 및 기타 서비스	23	0.2%	31	0.2%	30	0.2%		
계	14,619	100.0%	15,935	100.0%	19,750	100.0%		

Note: 기관 수는 조사표 회수 기업체 중 해당 연도에 연구개발비가 있는 기업체 수임.

<표 2-6>은 2011년도 조사에서 응답 기업체의 산업대분류 및 규모별 현황을 정리한 것이다. 앞서 살펴본 바와 같이 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’의 비중이 높고, 100인 미만의 기업체 규모에서 높은 비중을 차지하고 있다.

참고로 <부록 1>에는 산업중분류 및 규모별 응답 기업체 현황을 정리하여 수록하였다. 응답한 기업체가 있는 전체 63개 산업중분류 중에서 12개 산업중분류에서 해당 기업체 수가 500개소를 넘었고, 이들 산업중분류에 속한 전체 기업체는 14,148개소(71.6%)로 일부 산업중분류에 집중되어 나타나고 있음을 알 수 있다.

<표 2-6> 산업대분류 및 규모별 기업체 현황: 2011년도

산업대분류	규모	종업원 수 기준						
		10인 미만	10-49인	50-99인	100-299인	300-999인	1000인+	합계
A. 농업, 임업 및 어업		9	16	2	5	1	1	34
B. 광업		0	0	0	0	0	0	0
C. 제조업		2,003	7,308	1,942	1,530	420	150	13,353
D. 전기, 가스, 증기 및 수도사업		10	14	1	1	3	6	35
E. 하수·폐기물처리, 원료재생 및 환경복원		15	75	16	2	2	1	111
F. 건설업		138	1,231	145	65	28	19	1,626
G. 도매 및 소매업		120	481	95	42	12	5	755
H. 운수업		0	8	0	4	2	3	17
I. 숙박 및 음식점업		0	3	0	1	1	2	7
J. 출판, 영상, 방송통신 및 정보서비스		378	1,377	283	172	32	16	2,258
K. 금융 및 보험업		0	1	0	1	0	0	2
L. 부동산업 및 임대업		2	5	2	1	1	0	11
M. 전문, 과학 및 기술서비스업		270	743	210	138	40	14	1,415
N. 사업시설관리 및 사업지원 서비스업		2	21	8	7	4	7	49
P. 교육 서비스업		3	21	4	6	3	0	37
Q. 보건업 및 사회복지 서비스업		0	1	0	0	0	0	1
R. 예술, 스포츠 및 여가관련 서비스업		4	4	0	0	0	1	9
S. 협회 및 단체, 수리 및 기타 서비스		3	14	4	5	4	0	30
계		2,957	11,323	2,712	1,980	553	225	19,750

### 3. 연구개발비 및 연구개발인력 현황 분석

#### (1) 연구개발비 현황

2009년-2011년도 「연구개발현황조사」에서 연구개발비 현황을 분석한다. <표 2-7>은 연구주체별 연구개발비 현황이다. 전체 연구개발비 중 기업체의 비중이 약 75%로 가장 높고, 약간씩 높아지는 추세로 보인다. 이후로 연구개발비 총액은 공공연구기관, 대학, 병원 등의 순이고, 연도별 비중은 큰 변화가 없다.

<표 2-7> 연도 및 연구주체별 연구개발비 현황 (단위: 억원)

구분	2009년도		2010년도		2011년도	
	연구개발비	열%	연구개발비	열%	연구개발비	열%
공공연구기관	55,384	14.6%	62,568	14.3%	66,198	13.3%
대학	40,726	10.7%	45,653	10.4%	48,345	9.7%
의료기관	1,704	0.4%	2,329	0.5%	2,573	0.5%
기업체	281,470	74.2%	327,999	74.8%	381,787	76.5%
전체	379,285	100.0%	438,548	100.0%	498,904	100.0%

<표 2-8>은 최근 3개년 조사에 대해서 기업체의 규모별 연구개발비 현황을 정리한 것이다. 종업원 수 기준으로 1000명 이상의 대기업(약 220개소)의 연구개발비가 전체 기업체 연구개발비의 60%를 차지하고 있다. 반면 100인 미만의 기업체의 연구개발비 비중은 약 16%를 차지하고 있는데, 응답 기업체 수 기준으로 보면 100인 미만의 기업체는 앞선 <표 2-4>에서 살펴본 바와 같이 전체 기업체의 약 85%로 대부분을 차지하고 있다.

<표 2-8> 연도 및 기업체 규모별 연구개발비 현황

구분	2009년도		2010년도		2011년도	
	연구개발비	열%	연구개발비	열%	연구개발비	열%
10인 미만	4,036	1.4%	3,734	1.1%	5,443	1.4%
10-49인	30,639	10.9%	33,385	10.2%	41,992	11.0%
50-99인	16,511	5.9%	16,616	5.1%	19,197	5.0%
100-299인	25,729	9.1%	26,373	8.0%	30,552	8.0%
300-999인	25,289	9.0%	26,676	8.1%	29,878	7.8%
1000인+	179,268	63.7%	221,215	67.4%	254,724	66.7%
총합	281,470	100.0%	327,999	100.0%	381,787	100.0%

<표 2-9>는 기업체의 산업대분류별 연구개발비 현황을 정리한 것이다. 기업체 대상의 전체 연구개발비 중 ‘제조업’이 차지하는 비중은 약 87%를 차지하고 있고, ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’의 비중이 상대적으로 높다. 나머지 산업대분류가 차지하는 연구개발비 비중은 미미한 수준이다.

<표 2-9> 연도 및 산업대분류별 연구개발비 현황 (단위: 억원)

연도 산업대분류	2009년도		2010년도		2011년도	
	연구개발비	열%	연구개발비	열%	연구개발비	열%
A. 농업, 임업 및 어업	0	0.0%	260	0.1%	581	0.2%
B. 광업	0	0.0%	188	0.1%	0	0.0%
C. 제조업	243,549	86.5%	287,373	87.6%	334,254	87.5%
D. 전기, 가스, 증기 및 수도사업	2,723	1.0%	2,940	0.9%	3,795	1.0%
E. 하수·폐기물처리, 원료재생 및 환경복원	206	0.1%	203	0.1%	316	0.1%
F. 건설업	8,460	3.0%	7,455	2.3%	9,085	2.4%
G. 도매 및 소매업	4,644	1.6%	5,395	1.6%	6,152	1.6%
H. 운수업	480	0.2%	742	0.2%	1,222	0.3%
I. 숙박 및 음식점업	81	0.0%	8	0.0%	66	0.0%
J. 출판, 영상, 방송통신 및 정보서비스	13,067	4.6%	15,358	4.7%	16,909	4.4%
K. 금융 및 보험업	17	0.0%	13	0.0%	10	0.0%
L. 부동산업 및 임대업	323	0.1%	224	0.1%	102	0.0%
M. 전문, 과학 및 기술 서비스업	7,256	2.6%	7,002	2.1%	8,385	2.2%
N. 사업시설관리 및 사업지원 서비스업	423	0.2%	566	0.2%	540	0.1%
P. 교육 서비스업	124	0.0%	129	0.0%	164	0.0%
Q. 보건업 및 사회복지 서비스업	2	0.0%	2	0.0%	2	0.0%
R. 예술, 스포츠 및 여가관련 서비스업	34	0.0%	34	0.0%	28	0.0%
S. 협회 및 단체, 수리 및 기타 서비스	80	0.0%	106	0.0%	176	0.0%
계	281,470	100.0%	327,999	100.0%	381,787	100.0%

(2) 연구개발인력 현황

2009년~2011년도 조사에 대한 연구개발인력 현황을 분석한다. <표 2-10>은 연구주체별 연구개발인력 현황이다. 우리나라의 연구개발인력은 매년 증가하고 있으며, 기업체의 비중이 약 55%로 가장 높고, 다음으로 대학의 비중이 약 35% 수준을 보이고 있다. 연구개발인력에 대한 연구주체별 구성비는 연도에 따라 큰 변화 없이 유지되고 있다.

<표 2-10> 연도 및 연구주체별 연구개발인력 현황 (단위: 인)

구분	2009년도		2010년도		2011년도	
	인원	열%	인원	열%	인원	열%
공공연구기관	36,311	7.8%	38,845	7.8%	42,823	8.1%
대학	171,223	36.7%	183,982	36.8%	181,724	34.2%
의료기관	3,897	0.8%	5,162	1.0%	5,707	1.1%
기업체	255,393	54.7%	272,135	54.4%	300,877	56.6%
전체	466,824	100.0%	500,124	100.0%	531,131	100.0%

<표 2-11>은 최근 3개년 조사에 대해서 기업체의 규모별 연구개발인력 현황을 정리한 것이다. 종업원 수 기준으로 1,000명 이상의 대기업(약 220개소)의 연구개발비가 전체 기업체 연구개발인력의 약 44%를 차지하고 있다. 100인 미만의 기업체의 연구개발인력 비중은 약 35%를 차지하고 있다. 연구개발인력의 기업체 규모별 구성비는 연도에 따라 큰 변화없이 유지되고 있다.

<표 2-11> 연도 및 기업체 규모별 연구개발인력 현황

구분	2009년도		2010년도		2011년도	
	인원	열%	인원	열%	인원	열%
10인 미만	8,064	3.2%	7,223	2.7%	10,320	3.4%
10-49인	53,877	21.1%	57,268	21.0%	67,157	22.3%
50-99인	23,740	9.3%	25,045	9.2%	27,577	9.2%
100-299인	35,172	13.8%	34,922	12.8%	38,497	12.8%
300-999인	25,351	9.9%	26,244	9.6%	25,976	8.6%
1000인+	109,189	42.8%	121,433	44.6%	131,350	43.7%
총합	255,393	100.0%	272,135	100.0%	300,877	100.0%

<표 2-12>는 기업체의 산업대분류별 연구개발인력 현황을 정리한 것이다. 기업체 대상의 전체 연구개발인력 중 '제조업'이 차지하는 비중은 약 78%를 나타내고 있고, '제조업', '출판, 영상, 방송통신 및 정보서비스업', '건설업', '전문 과학 및 기술서비스업'의 비중이 상대적으로 높다. 나머지 산업대분류가 차지하는 연구개발인력이 차지하는 비중은 낮은 수준이다.

<표 2-12> 연도 및 산업대분류별 연구개발인력 현황 (단위: 인)

연도 산업대분류	2009년도		2010년도		2011년도	
	인원	열%	인원	열%	인원	열%
A. 농업, 임업 및 어업	0	0.0%	305	0.1%	485	0.2%
B. 광업	0	0.0%	112	0.0%	0	0.0%
C. 제조업	199,604	78.2%	212,922	78.2%	236,472	78.6%
D. 전기, 가스, 증기 및 수도사업	1,169	0.5%	1,225	0.5%	1,303	0.4%
E. 하수·폐기물처리, 원료재생 및 환경복원	367	0.1%	340	0.1%	581	0.2%
F. 건설업	8,949	3.5%	9,010	3.3%	9,760	3.2%
G. 도매 및 소매업	6,497	2.5%	6,886	2.5%	7,378	2.5%
H. 운수업	451	0.2%	368	0.1%	452	0.2%
I. 숙박 및 음식점업	127	0.0%	27	0.0%	90	0.0%
J. 출판, 영상, 방송통신 및 정보서비스	25,426	10.0%	27,875	10.2%	29,849	9.9%
K. 금융 및 보험업	21	0.0%	21	0.0%	10	0.0%
L. 부동산업 및 임대업	384	0.2%	324	0.1%	126	0.0%
M. 전문, 과학 및 기술 서비스업	11,348	4.4%	11,484	4.2%	13,093	4.4%
N. 사업시설관리 및 사업지원 서비스업	610	0.2%	623	0.2%	640	0.2%
P. 교육 서비스업	244	0.1%	355	0.1%	359	0.1%
Q. 보건업 및 사회복지 서비스업	5	0.0%	5	0.0%	22	0.0%
R. 예술, 스포츠 및 여가관련 서비스업	43	0.0%	50	0.0%	47	0.0%
S. 협회 및 단체, 수리 및 기타 서비스	148	0.1%	203	0.1%	210	0.1%
계	255,393	100.0%	272,135	100.0%	300,877	100.0%

(3) 2011년도 조사에 대한 종합 분석

<표 2-13>부터 <표 2-15>는 각각 연구주체, 기업체 규모, 기업체의 산업대분류별로 살펴본 조사대상 기관 수, 조사표 회수 기관 수, 연구개발비, 연구개발인력 등의 현황을 2011년도 조사에 대하여 정리한 것이다. 연구개발비의 분포는 연구개발인력 분포에 비해서 소수 대규모 기업체나 기관의 영향력이 큰 것으로 분석되었다.

<표 2-13> 연구주체별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도

구분	응답 기관 수		연구개발비(단위: 억원)		연구개발인력(단위: 인)	
	기관 수	열%	연구개발비	열%	인원	열%
공공연구기관	246	1.2%	66,198	13.3%	42,823	8.1%
대학	312	1.5%	48,345	9.7%	181,724	34.2%
의료기관	64	0.3%	2,573	0.5%	5,707	1.1%
기업체	19,750	96.9%	381,787	76.5%	300,877	56.6%
전체	20,372	100.0%	498,904	100.0%	531,131	100.0%

<표 2-14> 기업체 규모별 조사대상 기관, 연구개발비, 연구개발인력 현황: 2011년도

구분	응답 기관 수		연구개발비(단위: 억원)		연구개발인력(단위: 인)	
	기관 수	열%	연구개발비	열%	인원	열%
10인 미만	2,957	15.0%	5,443	1.4%	10,320	3.4%
10-49인	11,323	57.3%	41,992	11.0%	67,157	22.3%
50-99인	2,712	13.7%	19,197	5.0%	27,577	9.2%
100-299인	1,980	10.0%	30,552	8.0%	38,497	12.8%
300-999인	553	2.8%	29,878	7.8%	25,976	8.6%
1000인+	225	1.1%	254,724	66.7%	131,350	43.7%
총합	19,750	100.0%	381,787	100.0%	300,877	100.0%

<표 2-15> 산업대분류별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도

산업대분류	연도		응답 기관 수		연구개발비 (단위: 억원)		연구개발인력 (단위: 인)	
	기관 수	열%	연구개발비	열%	인원	열%		
A. 농업, 임업 및 어업	34	0.2%	581	0.2%	485	0.2%		
B. 광업	0	0.0%	0	0.0%	0	0.0%		
C. 제조업	13,353	67.6%	334,254	87.5%	236,472	78.6%		
D. 전기, 가스, 증기 및 수도사업	35	0.2%	3,795	1.0%	1,303	0.4%		
E. 하수·폐기물처리, 원료재생 및 환경복원	111	0.6%	316	0.1%	581	0.2%		
F. 건설업	1,626	8.2%	9,085	2.4%	9,760	3.2%		
G. 도매 및 소매업	755	3.8%	6,152	1.6%	7,378	2.5%		
H. 운수업	17	0.1%	1,222	0.3%	452	0.2%		
I. 숙박 및 음식점업	7	0.0%	66	0.0%	90	0.0%		
J. 출판, 영상, 방송통신 및 정보서비스	2,258	11.4%	16,909	4.4%	29,849	9.9%		
K. 금융 및 보험업	2	0.0%	10	0.0%	10	0.0%		
L. 부동산업 및 임대업	11	0.1%	102	0.0%	126	0.0%		
M. 전문, 과학 및 기술 서비스업	1,415	7.2%	8,385	2.2%	13,093	4.4%		
N. 사업시설관리 및 사업지원 서비스업	49	0.2%	540	0.1%	640	0.2%		
P. 교육 서비스업	37	0.2%	164	0.0%	359	0.1%		
Q. 보건업 및 사회복지 서비스업	1	0.0%	2	0.0%	22	0.0%		
R. 예술, 스포츠 및 여가관련 서비스업	9	0.0%	28	0.0%	47	0.0%		
S. 협회 및 단체, 수리 및 기타 서비스	30	0.2%	176	0.0%	210	0.1%		
계	19,750	100.0%	381,787	100.0%	300,877	100.0%		

#### 4. 기업체의 연구개발비 및 연구개발인력 현황 분석

「연구개발활동조사」에서 기업체가 차지하는 비중은 응답 기관 수 측면에서 보면 약 96%, 연구개발비는 약 76%, 연구개발인력은 약 56%를 차지하고 있다. 따라서 「연구개발활동조사」에 대한 표본조사 전환의 타당성과 효과성을 검토하고자 할 때 기업체에 대한 세부 분석은 반드시 필요하다. 여기서는 기업체 대상 조사결과를 심층 분석하여 표본조사로 조사방식을 전환하는 것에 대한 타당성을 살펴보고자 한다.

<표 2-16>은 전체 연구기관에 대한 연구개발비에 대한 분석 결과이다. 2011년도 조사에서 중앙값(2.6)에 비해 평균(24.5)로 9.6배 큰 값이며 왜도(skewness)가 119.2로 분포가 오른쪽으로 꼬리가 긴 심한 비대칭적 형태임을 알 수 있다. <표 2-17>은 기업체에 대한 연구개발비에 대한 분석 결과이다. 2011년도 조사에서 중앙값(2.5)에 비해 평균(19.3)로 약 8배 큰 값이며 왜도(skewness)가 121.6으로 오른쪽으로 꼬리가 긴 심한 비대칭적 분포 형태로 전체 기업체에 대한 분석 결과와 유사하다.

<표 2-18>과 <표 2-19>는 각각 전체 연구기관과 기업체에 대한 연구개발인력에 대한 분석 결과이다. 먼저 전체 연구기관의 연구개발인력에 대한 분포적 특성으로 중앙값은 5이고 평균은 26.1명이며 왜도는 78.2이다. 연구개발인력의 분포는 연구개발비 분포보다 심하지는 않지만 오른쪽으로 긴 꼬리를 갖는 비대칭적 형태로 나타나고 있다. 기업체 대상의 분석결과도 유사하게 나타나고 있다.

<표 2-16> 연도별 연구개발비 분포 현황: 전체 연구기관 (단위: 억원)

구분	연구개발비 분포		
	2009년도	2010년도	2011년도
기관 수	15,237	16,550	20,372
평균	24.9	26.5	24.5
표준편차	612.5	715.1	726.9
<b>왜도</b>	<b>101.9</b>	<b>106.9</b>	<b>119.2</b>
변동계수	2,460.5	2,698.6	2,968.2
최대값	70,425.5	85,895.8	97,085.4
99%	284.5	308.7	259.5
95%	39.6	35.8	32.5
90%	17.9	17.1	15.1
75% Q3	6.8	6.5	5.8
중위수	3.0	2.8	2.6
25% Q1	1.5	1.4	1.3
10%	0.7	0.7	0.6
5%	0.4	0.4	0.4
1%	0.1	0.1	0.1
최소값	0.0	0.0	0.0

<표 2-17> 연도별 연구개발비 분포 현황: 기업체 대상 (단위: 억원)

구분	연구개발비 분포(기업체)		
	2009년도	2010년도	2011년도
기업체 수	14,619	15,935	19,750
평균	19.3	20.6	19.3
표준편차	614.6	718.2	729.4
<b>왜도</b>	<b>104.8</b>	<b>109.4</b>	<b>121.6</b>
변동계수	3,192.2	3,489.0	3,773.0
최대값	70,425.5	85,895.8	97,085.4
99%	151.2	143.6	128.8
95%	28.9	27.7	25.1
90%	15.3	14.9	13.4
75% Q3	6.5	6.1	5.5
중위수	2.9	2.7	2.5
25% Q1	1.5	1.4	1.3
10%	0.7	0.7	0.6
5%	0.4	0.4	0.4
1%	0.1	0.1	0.1
최소값	0.0	0.0	0.0

<표 2-18> 연도별 연구개발인력 분포 현황 (단위: 인)

구분	연구개발비 분포		
	2009년도	2010년도	2011년도
기관 수	15,237	16,550	20,372
평균	30.6	30.2	26.1
표준편차	367.8	389.8	370.9
<b>왜도</b>	<b>65.6</b>	<b>69.7</b>	<b>78.3</b>
변동계수	1,200.6	1,289.8	1,422.5
최대값	35,636	39,688	42,100
99%	417	386	300
95%	53	48	40
90%	25	23	20
75% Q3	11	10	9
중위수	6	6	5
25% Q1	5	4	4
10%	3	2	2
5%	2	2	2
1%	1	1	1
최소값	0	0	0

<표 2-19> 연도별 연구개발인력 분포 현황: 기업체 대상 (단위: 인)

구분	연구개발비 분포(기업체)		
	2009년도	2010년도	2011년도
기업체 수	14,619	15,935	19,750
평균	17.5	17.1	15.2
표준편차	318.9	342.7	326.5
<b>왜도</b>	<b>98.5</b>	<b>101.4</b>	<b>112.6</b>
변동계수	1,825.2	2,006.7	2,143.4
최대값	35,636.0	39,688.0	42,100.0
99%	123.0	121.0	105.0
95%	35.0	32.0	28.0
90%	20.0	19.0	16.5
75% Q3	10.0	9.0	8.0
중위수	6.0	6.0	5.0
25% Q1	4.0	4.0	4.0
10%	3.0	2.0	2.0
5%	2.0	2.0	2.0
1%	1.0	1.0	1.0
최소값	1.0	1.0	1.0

<표 2-20>과 <표 2-21>은 각각 기업체 규모별 평균 연구개발비 및 연구개발인력 현황이다. 규모가 커질수록 연구개발비와 개발인력이 급격히 증가하고 있음을 알 수 있다.

<표 2-22>는 각 규모에서 구한 연구개발비와 연구개발인력의 상관계수 현황이다. 전체 상관계수가 0.99보다 클 정도로 두 변수 사이의 상관성은 대단히 높다. 기업체의 규모 구분에서 연구개발비와 연구개발인력의 상관계수도 기업체 규모가 커짐에 따라 점점 커지고 있다.

연구개발비나 연구개발인력의 분포는 오른쪽으로 긴 꼬리를 갖는 비대칭적인 분포를 갖고 몇몇 기관이 차지하는 비중이 대단히 높다. 이와 같은 분포적 특성을 고려할 때 기업체 대상의 연구개발비 조사에서는 심한 비대칭적 분포를 갖는 모집단에서 표본의 대표성을 제고할 수 있는 표본추출법인 전수층을 설정한 층화임의추출법이나 Lavalée & Hidiroglou (1988) 방법에 기초한 변형질사추출법(Modified cutoff sampling)의 적용을 고려할 수 있을 것이다.

<표 2-20> 기업체 규모별 평균 연구개발비 현황 (단위: 억원)

구분	2009년도		2010년도		2011년도	
	평균	표준편차	평균	표준편차	평균	표준편차
10인 미만	2.0	2.3	1.8	2.2	1.8	2.3
10-49인	3.8	5.4	3.7	4.9	3.7	5.8
50-99인	7.8	14.0	7.3	9.5	7.1	12.2
100-299인	15.0	23.2	14.8	21.4	15.4	23.7
300-999인	46.4	70.9	50.7	89.0	54.0	98.9
1000인+	841.6	5,033.1	1,001.0	6,028.7	1,132.1	6,753.1
전체	19.3	614.6	20.6	718.2	19.3	729.4

<표 2-21> 기업체 규모별 평균 연구개발인력 현황 (단위: 인)

구분	2009년도		2010년도		2011년도	
	평균	표준편차	평균	표준편차	평균	표준편차
10인 미만	4.1	2.1	3.4	1.6	3.5	1.8
10-49인	6.7	4.6	6.4	4.3	5.9	4.2
50-99인	11.2	9.7	11.0	9.2	10.2	9.1
100-299인	20.5	23.5	19.5	22.0	19.4	24.2
300-999인	46.5	53.7	49.9	57.9	47.0	55.1
1000인+	512.6	2,596.5	549.5	2,863.3	583.8	3,008.6
전체	19.3	614.6	20.6	718.2	19.3	729.4

<표 2-22> 기업체 규모별 연구개발비와 연구개발인력의 상관계수

구분	2009년도	2010년	2011년도
10인 미만	0.3659	0.3328	0.3728
10-49인	0.4098	0.4487	0.4123
50-99인	0.3953	0.5906	0.5048
100-299인	0.6731	0.6931	0.6566
300-999인	0.6941	0.6743	0.6205
1000인+	0.9955	0.9943	0.9936
전체	0.9947	0.9936	0.9930

## 5. 조사 개선을 위한 시사점

「연구개발활동조사」의 대상은 전국의 공공연구기관, 대학, 의료기관, 기업체 등이다. 최근 조사 데이터에 대한 분석을 통해서 조사 개선을 위한 다음과 같은 시사점을 얻을 수 있었다.

첫째, 전반적으로 조사대상 기관 및 회수된 기관은 매년 증가하고 있고, 특히 민간 기업체의 증가폭이 가장 커서 증가분의 대부분을 차지하고 있다. 반면 공공연구기관, 대학, 병원 등은 다소 감소하거나 변화가 없는 것으로 나타나고 있다. 전체 조사대상 기관과 응답 기관에서 기업체가 차지하는 비중은 90% 이상으로 절대적으로 높은 상황이다. 이에 따라 매년 지속적으로 조사대상이 늘어나게 되어 조사 관리와 비용 등이 급속히 증가함에 따라 표본조사로 전환해야 할 필요성이 있다.

둘째, 조사 회수율 현황은 전반적으로 약간씩 떨어지는 추세인 것으로 보인다. 공공연구기관, 대학, 의료기관 등의 회수율은 전반적으로 97%를 상회하고 있지만, 기업체의 회수율은 80% 초반의 값을 보이다가 2011년도 조사에서는 77.9%를 나타내고 있다. 전반적으로 기업체의 회수율은 공공연구기관, 대학, 의료기관과 비교하여 상당히 낮은 수준을 보이고 있다.

셋째, 「연구개발활동조사」에서 기업체의 연구개발비와 연구개발인력 측면에서 차지하는 비중은 각각 약 75%과 약 55%를 보이고 있다. 기업체 규모에 따라 연구개발비와 연구개발인력에 큰 차이가 있어, 두 조사항목에 대해서 오른쪽으로 꼬리가 긴 비대칭적 분포를 갖는다. 또한 몇 개의 대규모 기업체가 연구개발비와 연구개발인력 측면에서 차지하는 비중이 대단히 크다. 따라서 이 조사에서 기업체를 대상으로 표본조사를 적용한다면 전수층을 설정한 층화임의추출법을 적용하는 방안을 검토해야 할 것이다.

### 제 3 장 모집단 분석

본 장에서는 「연구개발활동조사」에 대한 모집단 중 기업체 모집단에 대한 특성을 분석한다. 앞서 살펴본 바와 같이 「연구개발활동조사」의 조사대상 중 기업체 비중은 90% 이상이고, 표본조사로 전환한다고 할 때 우선적으로 검토해야 할 대상은 기업체 모집단이다. 여기서는 2012년 12월 말 기준으로 작성된 「연구개발활동조사」 기업체 모집단을 분석한다.

「연구개발활동조사」의 기업체 대상 목표모집단은 조사 기준년도에 연구개발 활동을 시행한 전체 기업체이다. 이 조사의 표본추출틀은 법적 근거(기초연구진흥 및 기술개발 지원에 관한 법률 제14조, 동법 시행령 제16조, 미래창조과학부 고시 제2013-94호)에 의해서 등록된 기업부설연구소 및 연구개발 전담부서 현황 자료이다. 이 자료는 기업부설연구소와 연구개발 전담부서를 보유한 기업의 리스트로 한국산업기술진흥협회에서 수집·관리하고 있다. 원론적으로 보면 현재의 표본추출틀은 목표모집단을 포괄하지 못하는 부분이 있을 수 있지만, 기업부설연구소와 연구개발 전담부서를 보유한 기업의 리스트를 토대로 세제 혜택 등이 적용된다는 점을 고려할 때 미포함 비율은 높지 않을 것으로 판단된다.

조성표 등(2009)은 우리나라에서 연구개발을 수행하는 거의 모든 기업체가 한국산업기술진흥협회에 등록되어 있기 때문에 한국산업기술진흥협회에 등록된 연구소 보유기업체를 모집단으로 간주하는 것이 효율적이라고 주장하였다. 반면 유승훈과 문혜선(2002)은 해외의 사례를 들어 우리나라의 연구개발현황을 정확히 집계하기 위해서는 모집단의 확대와 표본조사의 도입이 필요하다는 의견을 제시하였다.

<표 3-1>은 기업부설연구소와 연구개발 전담부서를 보유한 기업의 산업대분류 및 기업체 규모별 현황이다. 전체 모집단 기업체는 36,404개소이고, 기업체 규모별 현황을 보면 소기업이 18,951개소(51.1%)로 가장 많고, 다음으로 벤처기업이 9,885개소(27.2%), 중기업이 5,651개소(15.5%), 대규모 기업체는 1,890개소(5.2%)이다.

산업대분류별 현황을 보면 ‘제조업’에 속한 기업체가 25,235개소(69.3%)를 차지하고 있고, 그 다음으로 ‘출판, 영상, 방송통신 및 정보서비스업’이 4,741개소(13.0%), ‘건설업’이 2,502개소(6.9%), ‘전문 과학 및 기술서비스업’이 2,110개소(5.8%) 등의 비중이 상대적으로 높다. ‘제조업’과 ‘출판, 영상, 방송통신 및 정보서비스업’의 벤처기업 또는 소규모 기업체는 24,248개소로 전체 모집단 기업체의 66.6%를 차지하고 있다.

<표 3-2>는 기업부설연구소와 연구개발 전담부서를 보유한 기업의 지역 및 기업체 규모별 현황이다. 경기도에 소재한 기업체가 가장 많은 12,557개소(34.5%)로 가장 많

고, 서울에 소재한 기업체가 9,220개소(25.3%)이다.

<표 3-3>는 모집단 기업체의 업종 및 기업체 규모별 현황이다. ‘전기·전자’ 업종에 속한 기업체가 가장 많은 9,102개소(25.0%)로 가장 많고, ‘기계’ 업종의 기업체가 6,034개소(16.6%)이다.

<표 3-1> 산업대분류 및 기업체 규모별 현황

산업대분류	기업규모					
	연구원, 교원 창업기업	벤처기업	소기업	중기업	대기업	총합
A. 농업, 임업 및 어업	0	10	13	17	6	46
B. 광업	0	0	6	0	1	7
C. 제조업	15	6,957	13,319	3,475	1,469	25,235
D. 전기, 가스, 증기 및 수도사업	0	8	13	13	15	49
E. 하수·폐기물처리, 원료재생 및 환경복원	0	34	37	91	9	171
F. 건설업	1	183	1,989	221	108	2,502
G. 도매 및 소매업	0	183	485	612	21	1,301
H. 운수업	0	1	13	8	18	40
I. 숙박 및 음식점업	0	1	1	7	2	11
J. 출판, 영상, 방송통신 및 정보 서비스업	4	1,936	2,036	624	141	4,741
K. 금융 및 보험업	0	2	2	2	1	7
L. 부동산업 및 임대업	0	1	3	3	4	11
M. 전문, 과학 및 기술 서비스업	7	532	972	517	82	2,110
N. 사업시설관리 및 사업지원 서비스업	0	23	30	19	8	80
P. 교육 서비스업	0	12	12	18	2	44
Q. 보건업 및 사회복지 서비스업	0	2	13	8	1	24
R. 예술, 스포츠 및 여가관련 서비스업	0	0	4	5	0	9
S. 협회 및 단체, 수리 및 기타 서비스업	0	0	3	11	2	16
총합	27	9,885	18,951	5,651	1,890	36,404

<표 3-2> 지역 및 기업체 규모별 현황

시도	기업규모					
	연구원, 교원 창업기업	벤처기업	소기업	중기업	대기업	총합
서울	5	2,606	4,620	1,590	399	9,220
부산	1	382	707	271	45	1,406
대구	0	349	734	206	39	1,328
인천	0	449	1,063	340	84	1,936
광주	0	227	428	88	23	766
대전	10	512	532	97	86	1,237
울산	0	105	154	88	74	421
경기	8	3,492	6,921	1,525	611	12,557
강원	0	136	210	54	18	418
충북	1	247	578	235	80	1,141
충남	0	323	668	310	137	1,438
전북	1	185	359	106	42	693
전남	0	124	281	73	26	504
경북	1	331	638	266	105	1,341
경남	0	392	978	390	109	1,869
제주	0	24	80	11	3	118
해외	0	1	0	1	9	11
총합	27	9,885	18,951	5,651	1,890	36,404

<표 3-3> 업종 및 기업체 규모별 현황

시도	기업규모					
	연구원, 교원 창업기업	벤처기업	소기업	중기업	대기업	총합
건설엔지니어링	2	292	2,107	539	150	3,090
금속	2	348	1,083	319	115	1,867
기계	3	1,539	3,217	920	355	6,034
기타	4	1,015	3,027	814	174	5,034
생명과학	1	290	241	91	45	668
섬유	0	105	566	223	50	944
소재	1	272	497	168	58	996
식품	0	162	537	240	87	1,026
전기·전자	6	3,308	4,140	1,166	482	9,102
정보처리	4	1,609	1,834	546	78	4,071
화학	3	708	1,311	514	277	2,813
환경	1	237	391	111	19	759
총합	27	9,885	18,951	5,651	1,890	36,404

<표 3-4>는 모집단 기업체의 연구분야 코드 및 기업체 규모별 현황이다. ‘전기·전자’ 분야의 기업체가 가장 많은 9,102개소(23.9%)로 가장 많고, ‘기계’ 업종의 기업체가 6,221개소(17.1%)이다.

<표 3-5>는 기업체 규모 및 연구소 인정 연도별 현황이다. 전체 기업체의 50.6%인 18,410개소는 2010년 이후 연구소 인정을 받았고, 10,367개소(28.5%)는 2005년부터 2009년도 사이에 인정을 받았다. 전체 기업체 중 약 79%는 2005년도 이후 연구소 인정을 받았고, 기업체 규모별 현황을 보면 벤처기업이나 소규모 기업체일 때 최근(2010년-)에 연구소 인정을 받은 비율이 대단히 높았다.

<표 3-4> 연구분야 코드 및 기업체 규모별 현황

시도	기업규모					
	연구원, 교원 창업기업	벤처기업	소기업	중기업	대기업	총합
SW개발공급	2	1,694	1,442	592	116	3,846
건설	0	116	493	119	73	801
건설엔지니어링	0	33	1,208	127	30	1,398
경영컨설팅	0	9	16	0	0	25
공학(엔지니어링)	2	197	424	341	60	1,024
교육기관	0	4	3	5	0	12
금속	2	304	1,040	296	108	1,750
기계	4	1,567	3,336	956	358	6,221
기타	4	492	1,515	336	98	2,445
문화및사업서비스	0	21	46	17	2	86
산업디자인	1	232	1,143	302	33	1,711
생명과학	1	331	290	117	60	799
섬유	0	81	341	148	37	607
소매	0	2	0	2	0	4
소재	1	328	618	206	71	1,224
시장조사	0	0	1	0	0	1
식품	0	151	537	244	87	1,019
위생산업	0	0	2	0	0	2
의료및보건	0	4	6	8	1	19
전기·전자	6	3,122	4,071	1,062	435	8,696
정보서비스	0	141	139	52	8	340
정보처리	0	93	476	68	10	647
화학	3	688	1,326	512	278	2,807
환경	1	275	478	141	25	920
총합	27	9,885	18,951	5,651	1,890	36,404

<표 3-5> 기업체 규모 및 연구소 인정 연도별 현황

구분	연구소 인정 연도					총합
	2010-2012년	2005-2009년	2000-2004년	1990-1999년	1990년 이전	
연구원, 교원	20	3	3	1	0	27
창업기업	74.1%	11.1%	11.1%	3.7%	0.0%	100.0%
벤처기업	4,670	3,184	1,334	678	19	9,885
	47.2%	32.2%	13.5%	6.9%	0.2%	100.0%
소규모	11,520	5,268	1,603	532	28	18,951
	60.8%	27.8%	8.5%	2.8%	0.1%	100.0%
중규모	1,847	1,628	1,202	847	127	5,651
	32.7%	28.8%	21.3%	15.0%	2.2%	100.0%
대규모	353	284	375	565	313	1,890
	18.7%	15.0%	19.8%	29.9%	16.6%	100.0%
총합	18,410	10,367	4,517	2,623	487	36,404
	50.6%	28.5%	12.4%	7.2%	1.3%	100.0%

## 제 4 장 현행 조사 개선 방안

### 1. 모집단 구성

「연구개발활동조사」의 조사대상 기관 및 회수된 기관은 매년 증가하고 있고, 특히 민간 기업체의 증가폭이 가장 커서 증가분의 대부분을 차지하고 있다. 전체 조사대상 기관과 응답 기관에서 기업체가 차지하는 비중은 90% 이상으로 절대적으로 높은 상황이다. 따라서 기업체 모집단을 잘 구축하여 관리하는 것은 조사의 정확도를 높이기 위해서 우선적으로 살펴볼 중요한 요소이다.

이 조사에서 기업체 대상 목표모집단은 조사 기준년도에 연구개발 활동을 수행한 전체 기업체이다. 이 조사를 위한 표본추출틀은 법적 근거(기초연구진흥 및 기술개발지원에 관한 법률 제14조, 동법 시행령 제16조, 미래창조과학부 고시 제2013-94호)에 의해서 인증된 「기업부설연구소 및 연구개발 전담부서 현황」자료이다. 이와 같이 구성된 표본추출틀은 원칙적으로 목표모집단을 포괄하지 못하는 부분이 있을 수 있지만, 미포함(undercoverage) 비율은 높지 않을 것이라는 점은 앞선 제3장에서 밝힌 바 있다. 여기서는 현행 모집단 리스트 상에서 관리되고 있는 조사항목을 살펴보고 표본조사로 전환될 때 추가되어야 할 조사항목에 대해서 살펴본다.

현행 모집단 리스트 상의 정보는 기업명, 기업규모, 연구분야 코드, 설립일자, 소재지, 업종코드, 연구소명, 산업분류 등 179개 항목에 이른다. 하지만 이들 조사항목 중 대부분은 법률에 의거해 인정된 기업부설연구소와 연구전담부서에 대한 관리 목적에 적합한 조사항목이다.

표본조사를 진행하기 위해서는 표본설계가 필수적이다. 효과적인 표본설계를 위해서는 표본추출틀이 전체 모집단을 잘 포괄해야 하고, 아울러 표본설계에 필요한 정보를 포함하고 있어야 한다. 현행 표본추출틀의 역할을 하고 있는 「기업부설연구소 및 연구개발 전담부서 현황」자료는 조사단위인 기업체에 대한 기본적인 정보가 부족한 상황이다. 예를 들어 기업체의 규모를 나타낼 때 현행의 표본추출틀은 ‘연구원, 교원 창업기업’, ‘벤처기업’, ‘소기업’, ‘중기업’, ‘대기업’ 등으로 구분하고 있다. 이 조사의 기업체 대상 표본설계에서 가장 기본적인 층화 정보는 산업분류, 기업체의 총 종업원 수, 기업체의 총 연구개발인력 등이다. 이 중 기업체의 총 종업원 수와 연구개발인력 등은 조사 기준시점에 따라 변화하게 된다. 따라서 조사기준 시점을 정해서 기업체의 총 종업원 수와 연구개발인력을 갱신하도록 하여 표본추출틀을 관리해야 한다. 아울러 해당 기업체가 ‘연구원, 교원 창업기업’ 또는 ‘벤처기업’인지 여부를 나타내는 조사항목을 추가하

여 관리한다면 현행의 표본추출틀 정보도 그대로 유지할 수 있다. 현재 표본추출틀에서는 기업체의 규모 구분을 ‘연구원, 교원 창업기업’, ‘벤처기업’, ‘소기업’, ‘중기업’, ‘대기업’ 등으로 구분하고 있어, 기업체의 총 종업원 수에 따라 구분된 것이라고 볼 수 없다.

조사기준 시점을 정해서 기업체의 총 종업원 수와 연구개발인력을 갱신하여 표본추출틀을 관리하는 방안으로는 직전 년도의 조사결과를 이용하여 활용하여 표본추출틀의 정보를 갱신할 수 있을 것이다. 이를 위해서는 사업체등록번호와 같은 각 기업체에 부여된 고유번호에 기초해서 조사데이터와 표본추출틀 정보를 연계하여 활용해야 할 것이다. 신규 기업체에 대해서는 해당 연도의 총 종업원 수와 연구개발인력을 입력하도록 한다면 추가적인 응답 부담 없이 표본추출틀의 관리가 가능할 것이다. 휴폐업 기업체에 대해서는 개별적인 확인 작업을 통해서 표본추출틀을 정비해야 할 것이다. 이와 같이 표본추출틀을 잘 관리하고 최신 정보로 갱신하는 것은 표본조사로 전환하고자 할 때 가장 기본적이고 중요한 작업이다.

현행 「연구개발활동조사」에서 조사표에 응답한 기업체 중에서 연구개발비와 연구개발인력이 모두 0이라고 응답한 기업인 경우도 조사데이터에 이를 포함해야 하고, 표본추출틀 정보에도 반영되어야 한다.

우리나라의 경우 「연구개발활동조사」에서 표본추출틀로 활용되고 있는 「기업부설 연구소 및 연구개발 전담부서 현황」 자료가 목표모집단을 대체로 잘 포괄하여 미포함(undercoverage) 비율이 높지 않을 것으로 예상되지만, 선진국의 경우는 전체 기업체를 대상으로 표본추출틀을 구성하여 전수조사와 표본조사를 병행하고 있다.

미국에서는 종업원 5인 이상을 고용하고 있는 185만개 기업을 대상으로 연구개발비(R&D 지출액)를 기준으로 전수조사와 표본조사를 병행하여 32,084개 기업을 추출하여 조사하였다. 일본은 「사업소·기업 통계조사」 중 자본금 1천만엔 이상인 기업 약 53만개를 대상으로 자본금 10억엔 이상의 대기업은 전수조사를 하고, 자본규모와 산업을 기준으로 기업을 선정하여 표본조사를 실시하고 있다(조성표 등, 2009).

우리나라의 경우도 「기업부설연구소 및 연구개발 전담부서 현황」 자료에 포함되지 않는 기업체들을 대상으로 일부 표본을 추출하여 「연구개발활동조사」를 진행함으로써 현행 표본추출틀의 과소포함으로 인한 연구개발비 과소 추정 여부에 대해서 살펴봐야 할 것이다.

## 2. 특이치 검출 및 처리 방안

### (1) 개요

기업체 대상의 표본조사에서 특이치의 문제는 그 동안 추정결과의 질을 좌우하는 중

요한 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 추정 과정에서 중요한 역할을 하게 되고, 특이치를 어떤 방법으로 검출하여 처리하였는가에 따라 통계의 정확도는 중대한 영향을 받게 된다.

일반적으로 표본조사에서의 특이치는 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. 「연구개발활동조사」가 전수조사로 진행될 때는 특이치가 발생하더라도 그 영향은 제한적이다. 하지만 이 조사가 기업체 대상의 표본조사로 바뀌게 되면 주요 관심모수가 모총계이기 때문에 조사값과 가중치의 곱의 형태( $\sum_{i \in \text{표본}} w_i \times y_i$ )로 추정과정에 반영됨으로써 특이치의 영향이 크게 나타나게 된다.

표본조사에 특이치 발생 원인을 정리하면 다음과 같다. 우선 특이치는 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.

특이치가 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. 만약 총계 추정이 목적이라면 추정량은 ' $\sum_{i \in \text{표본}} w_i \times y_i$ '이 될 것이기 때문에 ' $w_i \times y_i$ '이 매우 큰 값을 갖는 경우가 발생하게 된다.

특이치 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)을 참고하기 바란다.

특이치 발생의 다른 원인으로는 데이터 생성 과정의 오류(data capture error)를 들 수 있다. 데이터 생성 과정의 오류는 데이터 입력에서 오류가 발생하거나 응답과정에서 오류가 발생하여 매우 크거나 매우 작은 값이 입력되는 경우를 말한다. 이와 같은 특이치는 검출되면 값을 다시 조사하여 수정하거나 분석에서 제외하거나 결측 처리 후 대체방법 등을 이용하여 분석한다.

## (2) 특이치 검출 방법

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

### (가) 범위 에디팅 방법

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

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

이 방법은 일반 통계 분야에서 사용되는 가장 단순한 방법으로 조사값에 포함된 특이치로 인하여 표본표준편차가 커지게 되어 실제 특이치이지만 특이치로 검출되지 않는 매스킹 효과(masking effect) 현상이 발생할 수 있다. 매스킹 효과에 대한 대처방안으로 로버스트한 방법을 사용하여 중심위치와 산포의 측도를 산출하여 사용하는 방안을 고려할 수 있다. 예를 들어 표본평균 대신 중앙값을 사용하고, 표본표준편차대신 사분위수 범위나  $MAD = \text{median}|x - \text{median}(x)|$  을 사용한다.

### (나) 저항 울타리 방법

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

### (다) 비대칭적 저항 울타리 방법

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

#### (라) 로버스트 회귀모형 방법

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

#### (마) Hidiroglou and Berthelot 방법

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

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

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

단,  $r_{median}$  은  $r_i$ 의 중앙값이다.

$\textcircled{3} e_i = s_i \{ \text{MAX}(x_{i,t}, x_{i,t+1}) \}^u$  을 계산한다. 단  $\text{MAX}()$ 는 최대값을 계산하는 함수이고  $0 \leq u \leq 1$ 이며,  $u=1$ 이면 큰 조사값을 갖는 단위의 작은 변동에 더 큰 중요도를 부여한다는 의미이다.

$\textcircled{4} d_{D_L} = \text{MAX}(e_{median} - e_{q_1}, |a \cdot e_{median}|)$  와  $d_{D_U} = \text{MAX}(e_{q_3} - e_{median}, |a \cdot e_{median}|)$ 를 계산한다. 단  $e_{q_1}, e_{median}, e_{q_3}$ 는 각각 제1사분위수, 중앙값, 제3사분위수이며  $a = 0.05$ 가 추천된다.

(5)  $(e_{median} - c \times d_{D_L}, e_{median} + c \times d_{D_U})$ 의 구간 밖에 위치한 것을 특이치로 검출하게 된

다. 일반적으로  $c=6$ 이 사용된다.

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

이의규(2010)는 광업 및 제조업 조사에 Hidiroglou and Berthelot 방법을 사용하여 특이치를 검출하였고, 특이치의 식별범위를 극단이상치와 보통이상치로 나누어서 소개하였다.

### (3) 특이치 처리 방법

데이터 에디팅(data editing) 단계에서 검출된 특이치에 대해서는 우선 조사과정에서 아무런 문제없이 조사되고 정확히 기입된 것인지를 확인해야 한다. 이러한 점검 과정을 통해서 아무런 문제없는 조사값으로 확인된 특이치에 대해서는 추정 과정에서 적절하게 처리될 수 있도록 데이터 상에 표시되어야 하고, 대체(imputation) 과정에서 기증자(donor)로 사용되어서는 안 된다. 일반적으로 검출된 특이치에 대해서는 추정과정에서 특이치의 영향이 감소하도록 값 조정이나 가중치 조정의 과정을 거치게 된다. 이렇게 특이치의 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 특이치의 처리방법을 결정해야 한다.

#### (가) 가중값 조정(weight modification)

크기가  $N$ 인 모집단으로부터 단순임의추출법에 의해 뽑힌 표본 중 특이치에 해당하는 부분의 크기와 집합을 각각  $n_2$ ,  $s_2$ 라 하고, 특이치가 아닌 부분의 갯수와 집합을  $n_1$ ,  $s_1$ 로 정의하자. 여기서 특이치에 해당하는  $s_2$ 에 속한 조사단위의 가중치를 '1'로 낮추어 추정량을 계산하는 방안을 고려할 수 있다. 이 추정량은 특이치에 해당되지 않는 부분의 가중치가 보정되어 모집단의 총계는 다음 추정식에 의해서 추정된다.

$$\hat{Y}_{R1} = \sum_{s_1} \frac{N-n_2}{n_1} y_i + \sum_{s_2} y_i$$

이 추정량의 형태는  $s_2$ 를 전수층으로 구분한 절사표본추출법(cut-off sampling)에 의한 표본에서 구한 모총계 추정량의 형태와 같다. 절사표본추출법은 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 사업체 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 특이치 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 특이치 보정 요인  $f_i^{OAF}$ 은 두 부분으로 나누어 보정된다. 먼저 특이치에 해당하는 조사단위의 보정 요인은

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

이고, 특이치가 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 (이상은, 2008).

위의 가중치 작성과정에서 특이치 보정 과정은 특이치에 대해서 사실상 설계가중치를 1로 처리하는 것과 같다.

#### (나) 값 조정(value modification)

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

특이치로 검출된 조사값은 사전에 정해진 원칙에 따라 조정되는 데 일반적으로 널리 사용되고 있는 것은 원저화 방법이다.

#### (다) 가중값과 값의 동시 조정(weight and value modification)

가중값과 값을 동시에 조정하는 방법은 기본적으로 가중값 조정이나 값 조정을 동시에 고려함으로써 특이치의 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. 사업체조사에서는 흔히 가중치와 조사변수의 값의 변동으로 인해  $w_i y_i$ 의 값이 지나치게 커질 수 있다. 그러므로 가중치만

낮추어 조정하거나 또는 값만 조정하는 방법은 특이치를 완전히 조정할 수 없게 된다. 이러한 이유로 제안된 가중치와 값을 동시에 조정하는 방법은 다음과 같다 (Tambay, 1988).

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

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 특이치 검출의 기준값이고,  $s_2^-$ ,  $s_2^+$ 는 각각 특이치로 검출된 조사단위 중에서 특이치 검출 기준값인  $K_L$ 과  $K_U$ 보다 작거나 큰 경우의 조사단위를 의미한다.

#### (라) 로버스트 추정(Robust prediction)

모집단의 총계나 평균을 추정하는 경우의 로버스트 추정량은 특이치에 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 특이치의 영향을 덜 받는 모형기반 추정량(model-based estimator)을 이용하는 것이다. 이 방법은 전통적인 설계기반의 추정량을 이용하는 경우에는 거의 사용되지 않지만 스위스 연방통계청에서는 특이치 처리의 중요한 수단으로 사용하고 있다.

#### (4) 특이치 검출 및 처리 적용 방안

「연구개발활동조사」를 표본조사로 진행하게 되면 필연적으로 가중치를 사용해야 하고, 이에 따라 특이치가 발생할 경우에는 총계 추정에 미치는 영향이 대단히 클 수 있다. 따라서 조사데이터에 대한 특이치 검출과 처리는 추정의 정확도에 지대한 영향을 미칠 수 있다.

「연구개발활동조사」의 핵심 변수는 연구개발비와 연구개발인력이라고 할 수 있다. 이들 변수에 대한 특이치 검출 방법은 가능한 전년도 조사결과를 기초로 하여 Hidiroglou and Berthelot(1986) 방법을 적용하는 방안이 적합할 것이다. 이 방법은 당해 연도 조사결과와 전년도 조사결과와의 비(ratio)에 기초한 특이치 검출방법으로 자세한 내용은 앞서

다루었다. 추가로 연구개발인력과 연구개발비에 대해서 Hidioglou and Berthelot(1986) 방법을 적용하는 것도 고려할 수 있다. 이와 같은 특이치 검출 과정은 사전에 시스템적으로 마련되어 자동화할 수 있다. 이렇게 검출된 특이치에 대해서는 설계가중치를 1로 처리하여 추정하는 방안이 적용하기 간편하고 효과적인 방안이 될 것이다.

만약 현행의 조사방식과 마찬가지로 전수조사의 기본틀을 유지하여 조사를 진행한다고 하면 앞서 제시한 로버스트 추정방법에 기초한 모형기반 추정량(model-based estimator)을 이용하는 것도 고려할 수 있다.

참고로 신동완과 김재광(2006)은 당해연도의 연구개발비를 종속변수로 하고, 전년도 연구개발비, 매출액, 전년도 매출액, 총 연구개발인력, 전년도 총 연구개발인력 등의 총 5개 변수를 설명변수로 하는 ADL(Autoregressive Distributed Lag) 모형을 적합하여 상대예측오차(Relative Prediction Error)를 계산해 그 값이 클 때 특이치로 검출하여 조사된 연구개발비를 재검토하거나 재조사할 것을 제안하였다.

### 3. 조사데이터의 신뢰성 검토를 위한 재조사 방안

현행 「연구개발활동조사」는 우편 및 온라인을 통한 자계식(自計式) 조사를 기본으로 하고, 전화조사 등을 부가적으로 실시하고 있다. 또한 조사표 회수가 불가능한 경우는 휴·폐업이거나 연구개발비와 연구개발인력이 0인 경우로 간주하여 분석되고 있다. 하지만 무응답의 문제는 조사의 정확도를 좌우하는 중요한 문제로 무응답 처리에 대한 정확한 기준과 방법이 마련되어야 한다.

유승훈·문혜선(2002)와 신동완·김재광(2006)은 「연구개발활동조사」에서 무응답의 문제를 다루어 항목 무응답에 대해서 대체하는 방법을 제시하고 있다. 이들 연구는 주로 단순 무응답(응답거절, 연락안됨 등)에 대한 무응답 대체 방법으로 연구개발비나 연구개발인력 등이 전년도 조사결과와 높은 상관관계(자기상관관계)가 있다는 점에 착안하여 비보정(ratio adjustment)에 의한 무응답 대체(imputation) 방안을 제시하였다. 여기서는 조사의 측정오차와 무응답의 문제를 다른 관점에서 다루고자 한다.

우선 측정오차 문제를 살펴보기 위해서 조사표를 회수한 기업체를 대상으로 사후 표본 조사를 통해서 점검하는 방안을 살펴본다.

사후 표본조사 점검은 비표본오차를 측정하기 위해서 본 조사를 실시한 조사원들보다 더 잘 훈련되고 경험이 많은 조사원을 이용하여 표본조사에서 조사된 표본 중 부차표본을 추출하여 재조사하는 방법이다. 이 방법은 부차표본에 대한 재조사를 본 조사 후 곧바로 실행하여 가능한 한 재조사 오차를 줄여야만 효과적이다. 사후 표본조사 점검의 특별한 이점은 점검 표본에 포함된 단위들에 대한 두 개의 측정값을 대응시켜 관찰된 각각의 차이를 분석할 수 있다는 점이다. 만약에 대응된 두 개의 측정값이 상이하다면 비표본오차의

종류와 성질을 면밀히 분석하여 오차의 발생 원인을 살펴야 한다. 두 개의 자료 집합을 대응시키는 일은 시간과 비용의 문제로 어려움이 많지만 경우에 따라서 유용하게 이용될 수 있다. 만일 추출단위 간의 대응을 통한 관찰치 차이의 분석이 불가능할 때는 점검조사에서 얻어진 자료를 기초로 한 모수의 추정치들과 본 조사의 자료를 근거로 추정치들을 비교함으로써 순수한 오차에 대한 정보를 얻을 수 있을 것이다.

다음으로 무응답 기업체에 대한 부차표본추출 방안을 살펴본다. 첫 번째 조사를 실시할 때 조사대상 기업체 중에서 무응답 기업체가 발생하면 이들 무응답 기업체들 중에서 다시 부차표본을 추출하여 이 부차표본 내의 모든 기업체를 빠짐없이 조사하는 방법이다.

한센과 후르위츠(Hansen and Hurwitz, 1946)는 우편조사와 면접조사를 함께 사용하는 부차표본추출 방법을 제안하였다. 제안된 부차표본 기법은 첫 조사에서는 우편조사를 시행하고 재조사에서는 방문면접조사를 시행하는 방법이다. 이 방법을 이 조사에 적용하면 다음과 같다.

먼저  $n$ 개 기업체에게 설문지를 우송하여  $Rn$ 개 기업체로부터 조사표를 회수한 경우에 회수되지 않은  $(1-R)n$ 개 무응답 기업체 중에서  $(1-R)n/k$ 개소를 부차표본추출하여 개별방문 면접조사를 실시하는 방법이다. 이때 소요된 총비용은 다음과 같다.

$$C = C_1 Rn + C_2 \frac{(1-R)n}{k}$$

여기에서  $C_1$ 과  $C_2$ 는 각각 우편조사와 방문면접조사에서 기업체당 조사 비용과 자료 처리 비용을 포함한 금액이다. 이 경우 표본평균의 선형결합을 사용한 모평균의 추정량과 그의 분산은 다음과 같다.

$$\bar{y} = \frac{1}{n} \{ Rn\bar{y}_1 + (1-R)n\bar{y}_2 \}$$

$$V(\bar{y}) = R^2 S_r^2 / Rn + (1-R)^2 S_u^2 / \{(1-R)n/k\}$$

단,  $\bar{y}_1$ : 우편조사에서 응답 기업체들의 평균

$S_r^2$ : 응답 기업체 그룹의 분산

$\bar{y}_2$ : 부차표본추출하여 진행한 면접조사 응답 기업체들의 평균

$S_u^2$ : 무응답기업체 그룹의 분산

계산의 간편성을 위하여 우편조사 응답 그룹의 분산과 무응답 그룹의 분산이 서로 같다고(즉  $S_r^2/S_u^2=1$ ) 가정하면 무응답 기업체 중에서 추출되는 부차표본의 최적비율  $k$ 값은 다음과 같이 계산된다.

$$k = \left( \frac{C_2}{C_1} \right)^{1/2}$$

일반적으로 부차표본추출의 효과가 크기 위해서는  $k$ 가 5 또는 10이상이어야 한다.

이 조사에 한센과 후르위츠(Hansen and Hurwitz) 방법을 적용하기 위해서는 첫 번째 조사방법을 적용할 때의 조사단위당 비용과 방문면접조사를 진행할 때의 조사단위당 비용의 비를 알 수 있어야 하는데, 대략 전체 무응답 기업체 중 5~10% 정도를 표본추출하여 조사하는 방안을 적용할 수 있을 것이다.

## 제 5 장 표본조사 적용 방안

### 1. 표본조사 방식을 적용할 때 고려사항

「연구개발활동조사」의 조사대상 중 기업체가 차지하는 비중은 90% 이상으로 절대적이고 증가폭도 가장 커서 증가분의 대부분을 차지하고 있다. 따라서 기업체 모집단을 잘 구축하여 관리하는 것은 기업체 대상의 표본설계 과정에서 가장 기초적이고 우선적으로 고려해야 할 부분이다.

앞선 현행 조사에 대한 분석에서 살펴보았듯이 이 조사에 대한 회수율은 전반적으로 약간씩 떨어지는 추세인 것으로 보인다. 공공연구기관, 대학, 의료기관 등의 회수율은 전반적으로 97%를 상회하고 있지만, 기업체의 회수율은 80% 초반의 값을 보이다가 2011년도 조사에서는 77.9%를 나타내고 있다. 기업체의 회수율은 공공연구기관, 대학, 의료기관과 비교하여 상당히 낮은 수준을 보이고 있다. 따라서 표본조사를 적용한다면 기업체를 대상으로 우선적으로 적용하여 회수율을 높이기 위한 노력을 집중하는 것이 바람직할 것이다.

「연구개발활동조사」에서 연구개발비와 연구개발인력 측면에서 기업체의 비중은 전체의 약 75%와 약 55%로 공공연구기관, 대학, 의료기관 등에 비해서 월등히 높다. 또한 기업체 규모에 따라 연구개발비와 연구개발인력에 큰 차이가 있어, 두 조사항목에 대해서 오른쪽으로 꼬리가 긴 비대칭적 분포를 갖는다. 따라서 규모가 큰 기업체를 대상으로 전수층을 마련하여 조사를 진행하는 것이 추정의 정확도를 높일 수 있는 방안이 될 것이다. 따라서 본 연구에서는 100인 이상의 중규모 이상 기업체에 대해서는 전수추출하고, 100인 미만의 중소 규모 기업체에 대해서는 표본조사로 진행하는 방안을 모색한다. 좀더 체계적인 표본추출법으로 모집단의 분포적 특성을 고려하여 기업체 대상의 연구개발비 조사에서 표본의 대표성을 제고할 수 있는 표본추출법인 변형절사추출법(Modified cutoff sampling)을 적용하는 방안도 검토한다.

기업체의 산업대분류별 연구개발인력 현황을 살펴본 결과에 의하면 기업체 대상의 전체 연구개발인력 중 ‘제조업’이 차지하는 비중은 약 78%를 나타내고 있고, ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’의 비중이 상대적으로 높다. 나머지 산업대분류가 차지하는 연구개발인력이 차지하는 비중은 낮은 수준이다. <표 5-1>은 2011년도 조사결과를 분석한 것으로 산업대분류별 응답 기관

수, 연구개발비, 연구개발인력 현황이다. 앞서 제시한 4개 산업대분류를 제외한 나머지 산업대분류에 속한 기업체의 비중은 응답 기관 수의 5.6%, 연구개발비의 3.4%, 연구개발인력의 3.9%를 차지하고 있어 대단히 낮은 수준이라고 할 수 있다. 따라서 본 연구에서 고려하는 표본조사 대상은 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 100인 미만 기업체로 국한한다. 2011년도 조사에서 표본조사 대상인 100인 미만 기업체 수는 총 16,028개소로 전체 응답 기업체의 81.2%를 차지하고 있다.

현행 조사에서 통계작성 단위에는 산업대분류를 포함하고 있어, 전체 기업체 수가 작은 산업대분류에 대해서 전수조사를 진행하는 것은 불가피한 선택이다.

<표 5-1> 산업대분류별 응답 기관, 연구개발비, 연구개발인력 현황: 2011년도

산업대분류	구분		응답 기관 수		연구개발비 (단위: 억원)		연구개발인력 (단위: 인)	
	기관 수	열%	연구개발비	열%	인원	열%		
C. 제조업	13,353	67.6%	334,254	87.5%	236,472	78.6%		
F. 건설업	1,626	8.2%	9,085	2.4%	9,760	3.2%		
J. 출판, 영상, 방송통신 및 정보	2,258	11.4%	16,909	4.4%	29,849	9.9%		
M. 전문, 과학 및 기술 서비스업	1,415	7.2%	8,385	2.2%	13,093	4.4%		
기타 산업대분류	1,098	5.6%	13,154	3.4%	11,703	3.9%		
계	19,750	100.0%	381,787	100.0%	300,877	100.0%		

## 2. 모집단 분석: 표본조사 대상

앞서 살펴본 바와 같이 「연구개발활동조사」의 표본조사 대상을 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 100인 미만 소규모 기업체로 국한하였다. 현행 조사의 통계작성 단위 중 하나는 산업중분류 및 기업체 규모별 구분이다. 따라서 표본조사로 진행된다고 해도 산업중분류별 통계작성은 필수적이다.

<표 5-2>는 표본조사 대상 산업대분류에 대한 산업중분류별 2011년도 「연구개발활동조사」의 응답 기관 현황이다. 원칙적으로 보면 모집단 분석을 위해서는 표본추출틀의 역할을 하는 「기업부설연구소 및 연구개발 전담부서 현황」을 분석하는 것이 바람직하다. 하지만 현행의 표본추출틀은 기업체 규모를 나타낼 때 ‘연구원, 교원 창업기업’, ‘벤처기업’, ‘소기업’, ‘중기업’, ‘대기업’ 등으로 구분하고 있기 때문에 이를 직접적으로 이용할 수 없다. 반면 「연구개발활동조사」에서는 기업체의 종업원 수 정보를 포함하

고 있고, 이 조사가 원칙적으로 전수조사라는 점을 고려할 때 전체 모집단을 완벽하게 반영하고 있는 것은 아니지만 전체 모집단의 현황을 파악하는데 활용할 수 있다.

<표 5-2>를 보면 산업중분류 구분에 따라서 응답 기관수의 편차가 큼을 알 수 있다. 또한 현행 조사에서 산업중분류 및 기업체 규모별 통계작성이 이루어지고 있다는 점을 고려하면 일정 크기 이상의 기업체가 있는 산업중분류 및 기업체 규모 구분에서 표본조사를 적용하는 것이 통계 작성의 안정성 측면에서 바람직할 것이다. 본 연구에서는 적어도 200개소 이상의 기업체가 속해 있는 산업중분류 및 규모 구분에서 표본조사 방법을 적용할 것으로 제안한다. 이에 해당하는 산업중분류 및 규모에 대해서는 <표 5-2>에 음영을 넣어 표시하였으며, 표본조사 대상에 속한 기업체 수는 11,638개소이다.

<표 5-2> 산업중분류 및 기업체 규모별 조사대상 기업체 현황: 2011년도 조사

산업대분류	산업중분류	10인 미만	10-29인	30-99인	총합
C. 제조업	10	63	224	91	378
	11	6	15	9	30
	12	0	0	0	0
	13	30	146	66	242
	14	10	87	39	136
	15	5	29	6	40
	16	6	34	6	46
	17	6	54	16	76
	18	13	28	17	58
	19	5	16	5	26
	20	219	609	128	956
	21	30	99	35	164
	22	45	302	115	462
	23	31	151	39	221
	24	15	149	81	245
	25	88	540	122	750
	26	451	1,205	271	1,927
	27	216	565	124	905
	28	244	716	156	1,116
	29	382	1,616	359	2,357
F. 건설업	30	23	301	172	496
	31	15	89	40	144
	32	8	71	15	94
	33	92	262	30	384
	41	51	854	86	991
J. 출판, 영상, 방송통신 및 정보서비스	42	87	377	59	523
	58	213	892	192	1,297
	59	2	17	6	25
	60	0	1	1	2
	61	6	19	4	29
M. 전문, 과학 및 기술 서비스업	62	125	357	51	533
	63	32	91	29	152
	70	131	154	21	306
	71	9	39	11	59
총합	72	83	463	171	717
	73	47	87	7	141
	총합	2,789	10,659	2,580	16,028

### 3. 표본크기 결정

앞서 제시한 표본조사 대상의 산업중분류 및 기업체 규모 구분에 대하여 표본조사를 위한 표본크기를 결정한다. 앞서 밝힌 바와 같이 현행 조사분석 결과에서 통계작성 단위는 산업중분류 및 기업체 규모별 구분이다. 따라서 표본조사로 진행될 때 층화변수는 산업중분류 및 기업체 규모 구분이다.

표본크기 결정 단계에서는 각 산업중분류 및 기업체 규모 구분에서 총계 추정에 대한 목표 상대표준오차를 설정하여 이를 달성할 수 있도록 표본크기를 결정하였다. 각 층에서 표본크기는 다음 식에 따라 계산되었다.

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \quad n_0 = \left( \frac{\hat{C}}{D} \right)^2, \quad \hat{C} = S/\bar{y} \text{ (변동계수)}$$

본 연구에서 고려한 목표 상대표준오차 설정 방안과 표본크기는 <표 5-3>과 같다. 여기서 고려하는 주관심 변수는 연구개발비와 연구개발인력이다. 참고로 <부록 1>에는 표본크기 결정 단계에서 활용된 2011년도 조사에 대해서 산업중분류 및 기업체 규모별 응답 사업체의 연구개발비 총액 추정값, 연구개발비 표본 표준편차, 연구개발인력 총수, 연구개발인력 표본표준편차를 제시하였다.

본 연구에서 표본크기는 두 종류의 관심변수에 대해서 해당 상대표준오차를 모두 만족하도록 하였고, 각 산업중분류 및 기업체 규모 구분에서 안정적인 통계생산을 위해서 최소 표본크기를 200개소로 하여 구한 것이다. 예를 들어 방안 4를 최종 방안으로 결정할 경우에 필요한 표본크기는 8,668개소이고, 이에 따라 전수조사에 비해서 2,970개소를 줄여서 조사하게 된다. 이 방안을 적용할 때 각 산업중분류 및 기업체 규모 구분에서 연구개발비와 연구개발인력 총수 추정의 상대표준오차는 3% 이하를 만족하게 된다. 만약 통계작성 단위를 좀더 넓게 구성한다면 표본크기를 훨씬 더 줄일 수 있을 것이다.

<표 5-3> 목표 상대표준오차 설정 방안

기업체 수 기준	방안1	방안2	방안3	방안4
200-499개소	5%	5%	5%	3%
500-799개소	5%	5%	3%	3%
800개소 이상	5%	3%	3%	3%
표본크기	6,366	7,500	7,998	8,668
감소 표본수	5,272	4,138	3,640	2,970

<표 5-4>는 방안 4에 따라 표본조사를 진행할 경우에 각 층의 표본크기 현황을 전체 모집단의 크기와 비교하여 정리한 것이다. 이 표에서 전수층의 경우는 제외하여 정리한 것이다.

<표 5-4> 각 산업중분류 및 기업체 규모별 표본크기 현황: 방안 4

산업중분류	모집단 현황				표본조사 현황			
	10인 미만	10-29인	30-99인	총합	10인 미만	10-29인	30-99인	총합
10	63	224	91	378	63	200	91	354
20	219	609	128	956	202	443	128	773
22	45	302	115	462	45	282	115	442
25	88	540	122	750	88	480	122	690
26	451	1,205	271	1,927	316	729	247	1,292
27	216	565	124	905	200	444	124	768
28	244	716	156	1,116	216	500	156	872
29	382	1,616	359	2,357	317	949	295	1,561
30	23	301	172	496	23	291	172	486
33	92	262	30	384	92	235	30	357
41	51	854	86	991	51	578	86	715
42	87	377	59	523	87	292	59	438
58	213	892	192	1,297	200	609	192	1,001
62	125	357	51	533	125	297	51	473
72	83	463	171	717	83	346	171	600
전체	2,382	9,283	2,127	13,792	2,108	6,675	2,039	10,822

참고로 모집단 자료를 기초로 앞서 제시한 표본조사 방안을 적용하여 전체적으로 어떤 결과를 얻게 되는지 윤곽을 살펴보고자 한다. 현행 표본추출틀의 역할을 하고 있는 「기업부설연구소 및 연구개발 전담부서 현황」 자료는 기업체의 규모를 나타낼 때 ‘연구원, 교원 창업기업’, ‘벤처기업’, ‘소기업’, ‘중기업’, ‘대기업’ 등으로 구분하고 있다. 여기서는 앞선 분석과 마찬가지로 「연구개발활동조사」의 표본조사 대상을 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 ‘벤처기업’과 ‘소기업’으로 국한할 때 줄어드는 표본크기가 어느 정도 되는지 가늠해 보고자 한다.

<표 5-5>는 모집단 자료를 분석하여 얻은 것으로 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’에 속한 산업중분류의 ‘벤처기업’과 ‘소기업’에 대한 모집단 기업체 현황이다. 모집단 자료에 대한 전체 산업중분류 및 기업체 규모별 기업체 현황은 <부록 3>에 수록하였다.

<표 5-5>는 기업체 규모 구분을 ‘벤처기업’과 ‘소기업’ 등과 같이 넓게 구분하고 있기

때문에 이에 따라 표본조사의 대상이 되는 셀(산업중분류 및 규모)의 수가 늘어나게 된다<sup>1)</sup>. 즉, 현행의 표본추출틀 상의 정보를 이용해서는 현행 조사의 기업체 규모 구분(종업원 수 기준)을 적용하기 어렵지만, 기업체 규모 측면에서 통계작성 단위를 묶어서 설정하게 되면 표본크기를 훨씬 많이 줄일 수 있게 된다. 이와 같이 기업체 규모를 묶어서 「연구개발활동조사」 결과를 공표한다면 표본조사를 통해서 대략 10,000개소 이상의 표본을 줄일 수 있을 것으로 예상된다.

<표 5-5> 산업중분류 및 기업체 규모별 모집단 기업체 현황

산업대분류	산업중분류	연구원, 교원 창업기업	벤처기업	소규모	합계
C. 제조업	10	0	182	542	724
	11	0	15	28	43
	12	0	0	1	1
	13	0	93	354	447
	14	0	33	321	354
	15	0	19	92	111
	16	0	12	42	54
	17	0	27	110	137
	18	0	30	118	148
	19	0	6	24	30
	20	1	475	997	1,473
	21	1	139	145	285
	22	0	245	570	815
	23	0	96	302	398
	24	0	107	284	391
	25	4	402	1,055	1,461
	26	3	1,675	1,986	3,664
	27	1	686	821	1,508
	28	1	751	1,324	2,076
	29	3	1,441	2,773	4,217
30	0	239	634	873	
31	1	108	173	282	
32	0	31	143	174	
33	0	145	480	625	
F. 건설업	41	0	61	1,303	1,364
	42	1	122	686	809
J. 출판, 영상, 방송통신 및 정보서비스	58	2	1,221	1,178	2,401
	59	0	20	25	45
	60	0	2	3	5
	61	0	29	27	56
	62	2	551	680	1,233
63	0	113	123	236	
M. 전문, 과학 및 기술 서비스업	70	4	178	129	311
	71	0	30	78	108
	72	2	263	628	893
	73	1	61	137	199
	총합	27	9,608	18,316	27,951

자료: 「기업부설연구소 및 연구개발 전담부서 현황」, 2012년 12월 말 기준

1) 표본조사의 대상이 되는 해당 셀은 <표 5-5>에서 음영을 넣어 표시하였다.

참고로 이 조사와 같이 대규모 기업체의 비중이 커서 연구개발비나 연구개발인력의 분포가 심하게 오른쪽으로 긴 꼬리를 갖는 비대칭적 형태로 나타낼 때 적용될 수 있는 변형절사층화추출법에 대해서 살펴본다.

변형절사층화추출법(Modified cutoff sampling)은 Lavalée & Hidiroglou (1988) 방법에 기초한 것으로 Hidiroglou (1986)을 일반화한 알고리즘이다. 모집단의 왜도가 대단히 큰 모집단(오른쪽으로 꼬리가 긴 분포)에 대해서 적용되며, 목표상대표준오차와 표본층 (take-some stratum)에 대한 표본배분법(떡배분이나 네이만배분 등)이 정해진 상황에서 주어진 목표오차를 만족하면서 전체 표본크기를 최소화하는 층 경계점을 결정하는 것을 목적으로 한다.

Lavalée & Hidiroglou 방법은 전수층(take-all stratum)이 있는 층화임의추출법에서 표본크기 결정공식인 다음 식을 최소로 하는 층 경계점인  $b_{(1)}, b_{(2)}, \dots, b_{(L-1)}$  을 반복계산법 (iterative method)에 따라 구할 수 있는 알고리즘이다. 다음 식에서  $W_h, S_h^2, \bar{Y}_h$  등은 모두 층 경계점의 함수이다.

$$n = N \cdot W_L + \frac{N \left( \sum_{h=1}^{L-1} W_h^2 S_h^2 / a_h \right)}{N(cv_0 \bar{Y})^2 + \sum_{h=1}^{L-1} W_h S_h^2}$$

여기서,  $L$ 은 전체 층의 수,  $cv_0$ 는 목표상대표준오차,

$$a_h = \frac{N_h^p}{\sum_{h=1}^{L-1} N_h^p}, \quad (0 < p < \infty)$$

본 연구에서 산업분류만을 통계작성 단위로 하고, 기업체 규모 구분은 통계작성 단위로 고려하지 않는다면 변형절사층화추출법(Modified cutoff sampling)을 적용할 수 있다. 현행 조사에서는 산업중분류 및 기업체 규모별 추정도 통계작성 단위이기 때문에 전수층과 표본층을 구성하여 표본설계하였고, 엄밀한 의미의 변형절사층화추출법을 적용하지는 않았다. 하지만 100인 이상의 기업체에 대해서는 전수층을 구성하고, 100인 미만 기업체 규모에 대해서는 표본층을 구성했다는 점에서 변형절사층화추출법의 일종으로도 볼 수 있을 것이다.

#### 4. 표본추출 방안

본 연구에서는 「연구개발활동조사」의 표본조사 대상을 ‘제조업’, ‘출판, 영상, 방송 통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 100인 미만의 소규모 기업체로 국한하였고, 적어도 200개소 이상의 기업체가 속해 있는 산업중분류 및 구

모 구분에서 표본조사 방법을 적용할 것으로 제안하였다. 이와 같이 표본조사 대상을 가능한 축소할 것은 산업중분류 및 규모별 통계의 안정적인 생산이 가능하도록 하기 위함이다. 경우에 따라서 표본조사 대상을 좀더 넓혀 정의하게 되면 좀 더 표본크기를 줄일 수 있다.

표본조사로 진행되는 각 산업중분류 및 규모 층에서 표본사업체의 추출방법은 계통추출법을 적용한다. 각 세부 층에서는 해당 층에 속한 기업체를 소재지와 전년도 연구개발비를 기준으로 정렬한 후 표본크기만큼을 계통추출하여 표본기업체를 선정하여 조사한다.

향후 표본추출틀이 정비된 이후에 「연구개발활동조사」의 표본조사 기업체를 대상으로 휴폐업, 합병 및 신설 등에 따른 모집단 변화를 보다 체계적으로 적절히 반영하기 위해서는 영구난수를 이용한 표본추출법을 도입하는 것이 필요할 것이다. 일반적으로 영구난수(permanent random number, PRN)나 혹은 조정난수(collocated random number, CRN)는 표본추출틀 내 조사단위에 대해 0과 1사이의 난수를 부여하고 이를 계속되는 표본추출과정에서 영구적으로 사용하는 방법을 말한다. 이와 같이 영구난수법을 이용하여 표본을 추출하게 되면 반복되는 조사에서 매 조사별 표본간 중복 정도를 조절할 수 있어 표본 기업체의 응답부담이 집중되는 것을 막을 수 있게 된다.

영구난수를 사용하면 반복조사(repeated survey)시 표본간의 시점 간 연계(coordination over time)를 가능케 한다. 표본추출틀을 구성한 후 개별 개체에 난수를 부여하여 이를 영구적으로 사용하는데, 조사범위 내 개체들(persistants)는 기존 난수와 함께 남겨두고 생성된 개체(births)는 새롭게 독립적으로 발생시킨 난수와 함께 표본추출틀에 추가한다. 고정된 표본추출틀에 따라 표본을 선택함으로써 신설 기업체도 자연스럽게 사전에 정해진 표본추출틀에 따라 표본으로 추출되게 된다. 또한 영구난수를 활용하면 순환표본의 선택을 가능하게 한다. 매 시점 별 표본을 시작점  $a$ 에 일정한 상수  $\kappa$ 를 더해 줌으로 조사단위의 표본 포함을 순환시킬 수 있게 한다. 이때 상수  $\kappa$ 를 순환률(rotation rate)이라 칭한다.

영구난수를 이용한 표본추출법은 이와 같은 장점이 있기 때문에 선진국에서 기업체나 사업체 대상 조사에서 널리 사용되고 있다. 영구난수를 이용한 표본추출법에 대한 자세한 내용은 Srinath and Carpenter (1995)을 참조하기 바란다.

## 5. 추정 방법

모수 추정은 산업중분류 및 기업체 규모 구분에 따라 전수조사 대상과 표본조사 대상으로 구분하여 조사 업체를 선정하였기 때문에 전수조사 업체는 표본추출틀이 1이므로

표본추출에 따른 표집오차가 발생하지 않지만, 표본조사 업체는 계통추출법으로 선정하였으므로 동일한 업종의 표본조사 층 내에서는 모든 업체들이 동일한 추출확률을 갖게 된다.

전수층이 있는 층화계통추출법에서 모총계는 전수조사 층과 표본조사 층으로 구분하여 아래 식으로 추정한다.

$$\hat{Y} = Y^c + \sum_{h=1}^L \frac{N_h^s}{n_h^s} \sum_{k=1}^{n_h^s} y_{hk}^s$$

여기서  $L$ 은 표본층의 수이고,  $N_h^s$ 와  $n_h^s$ 는  $h$ 업종의 모집단 크기와 표본크기를 나타내며  $Y^c$ 는 전수층에 속한 업체들의 조사항목  $y$ 의 합계이다.

위의 식을 이용한 모평균의 추정은 모총계의 추정치를 전체 업체 수( $N$ )로 나누어 계산할 수 있으며 식으로 표현하면 아래와 같음

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N} = \frac{Y^c + \sum_{h=1}^L \frac{N_h^s}{n_h^s} \sum_{k=1}^{n_h^s} y_{hk}^s}{N}$$

모총계 추정량의 분산은 전수층으로부터 계산된 결과에서는 표본오차가 생기지 않고 표본층에서만 발생하므로 구해진 모총계 추정량의 분산은 아래 식으로 계산할 수 있다.

$$\begin{aligned} \widehat{Var}(\hat{Y}) &= \sum_{h=1}^L \widehat{Var}(\hat{Y}_h^s) \\ &= \sum_{h=1}^L \frac{(N_h^s)^2}{n_h^s(n_h^s - 1)} \sum_{k=1}^{n_h^s} (y_{hk}^s - \bar{y}_h^s)^2 \end{aligned}$$

여기서  $y_{hk}^s$ 는  $h$  산업중분류 및 규모 내의  $k$ 번째 표본사업체의 조사값,  $\bar{y}_h^s$ 는  $h$  산업중분류 및 규모의 표본업체들의 조사값의 평균이다. 계통추출법을 적용하여 표본업체를 선정했으나 업체들이 랜덤하게 분포되었다는 가정 하에 적용되는 단순임의추출법의 표본분산식을 이용하여 추정량의 분산을 계산한 결과이다.

위의 계산과정은 SPSS 또는 SAS 등의 복합표본조사 데이터 분석용 모듈을 이용하면 편리하게 계산할 수 있다.

## 제 6 장 결론 및 요약

### 1. 연구결과 요약

「연구개발활동조사」는 우리나라의 과학기술 연구 개발 활동(연구개발인력 및 연구개발비 등) 현황을 조사하여 국가 과학기술 정책 수립 등에 필요한 기초자료를 제공하고, 각계의 과학기술 분야 종사자로 하여금 과학기술 연구개발 계획 등에 참고자료로 제공하는 것을 목적으로 한다.

본 연구는 현행 「연구개발활동조사」의 조사 개선 방안을 제시하고, 이를 기초로 모집단 대표성이 확보될 수 있는 표본설계 방안을 제안하고, 이 방안의 타당성과 효과성을 검토하기 위함이다. 최근 3년 동안의 현행 조사 데이터에 대한 분석을 통해서 얻은 「연구개발활동조사」의 주요한 특징을 정리하면 다음과 같다.

첫째, 전반적으로 조사대상 기관 및 회수된 기관은 매년 증가하고 있고, 특히 기업체의 증가폭이 가장 커서 증가분의 대부분을 차지하고 있다. 반면 공공연구기관, 대학, 병원 등은 다소 감소하거나 변화가 없는 것으로 나타나고 있다. 전체 조사대상 기관과 응답 기관에서 기업체가 차지하는 비중은 90% 이상으로 절대적으로 높은 상황이다. 이에 따라 매년 지속적으로 조사대상이 증가하게 되어 조사 관리와 비용 등의 문제로 표본조사로 전환해야 할 필요성이 있다.

둘째, 조사 회수율 현황은 전반적으로 약간씩 떨어지는 추세로 나타났다. 공공연구기관, 대학, 의료기관 등의 회수율은 전반적으로 97%를 상회하고 있지만, 기업체의 회수율은 80% 초반의 값을 보이다가 2011년도 조사에서는 77.9%를 나타내고 있다. 전반적으로 기업체의 회수율은 공공연구기관, 대학, 의료기관과 비교하여 상당히 낮은 수준을 보이고 있다.

셋째, 「연구개발활동조사」에서 기업체의 연구개발비와 연구개발인력 측면에서 차지하는 비중은 각각 약 75%과 약 55%를 보이고 있다. 기업체 규모에 따라 연구개발비와 연구개발인력에 큰 차이가 있어, 두 조사항목에 대해서 오른쪽으로 꼬리가 긴 비대칭적 분포를 갖는다. 산업대분류 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’의 비중이 높고, 종업원 수 기준으로 1,000명 이상의 대기업(약 220개소)의 연구개발비가 전체 기업체 연구개발비의 60%, 전체 연구개발인력의 44%를 차지하고 있다. 따라서 기업체에 대해서 표본조사를 적용한다면 전수층을 설정한 층화임의추출법을 적용하는 것이 바람직하다.

넷째, 모집단에 속한 기업체는 36,404개소이고, 기업체 규모별 현황을 보면 소기업이 18,951개소(51.1%)과 벤처기업이 9,885개소(27.2%)에 집중되어 있다. 산업대분류별로 보면 ‘제조업’과 ‘출판, 영상, 방송통신 및 정보서비스업’의 벤처기업 또는 소규모 기업

체가 24,248개소로 전체 모집단 기업체의 66.6%를 차지하고 있다.

본 연구에서는 「연구개발활동조사」에 대해서 다음과 같은 조사 개선 방안을 제시하였다.

첫째, 현행 조사의 표본추출틀인 「기업부설연구소 및 연구개발 전담부서 현황」에서 조사단위인 기업체에 대한 기본적인 정보(총 종업원 수, 연구개발인력 등)를 보완하거나 추가하여 표본추출틀을 관리해야 한다. 또한 사업체등록번호와 같은 각 기업체에 부여된 고유번호에 기초해서 조사데이터와 표본추출틀 정보를 연계하여 활용해야 할 것이다. 표본추출틀을 잘 관리하고 최신 정보로 갱신하는 것은 표본조사로 전환하고자 할 때 가장 기본적이고 중요한 작업이다.

둘째, 기업체 조사에서 특이치와 가중치의 문제는 작성되는 통계의 정확도에 중대한 영향을 미치게 된다. 본 연구에서는 특이치 검출을 위한 여러 가지 방법을 제시하였고, 이들 중 Hidioglou and Berthelot(1986)에서 제시한 비(ratio)에 기초한 특이치 검출 방법과 가중치 조정방법을 제안하였다.

셋째, 이 조사에서는 조사표 회수가 불가능한 경우는 휴·폐업이거나 연구개발비와 연구개발인력이 0인 경우로 간주하여 분석되고 있다. 본 연구에서는 사후 표본조사 점검에 의한 조사 데이터의 신뢰성 검토 방안과 무응답 기업체에 대한 부차표본추출에 의한 조사방안을 제시하였다.

넷째, 이 조사의 표본조사 대상은 ‘제조업’, ‘출판, 영상, 방송통신 및 정보서비스업’, ‘건설업’, ‘전문 과학 및 기술서비스업’ 등의 100인 미만 소규모 기업체로 국한하는 것이 바람직하다. 본 연구에서는 표본조사 대상 모집단에 대하여 표본설계를 진행하고, 추정방법을 제시하였다.

## 2. 통계 작성 개선 로드맵

본 연구를 통해서 마련된 「연구개발활동조사」의 통계 작성 개선 과제를 구체적으로 실행하기 위해 요구되는 단계별 개선 전략을 정리하면 다음 <표 6-1>과 같다. 현재의 인적·물적 자원 및 작성환경 등을 고려할 때 통계 품질수준 제고를 위해 제시된 개선 과제들을 단기간에 모두 해결하는 것은 현실적으로 쉽지 않을 것으로 예상된다. 따라서 개선과제의 시급성 및 기대효과에 따라 단기(1년 이내), 중기(2년 이내) 및 장기(3년 이상) 과제로 분류하고 장기적인 개선 계획을 수립하여 단계별로 과제들을 해결하는 것이 효과적일 것이다.

<표 6-1> 「연구개발활동조사」 통계 작성 개선 로드맵

**「연구개발활동조사」 단계별 통계작성 개선 로드맵**

단기과제 (1년 이내)	
개선 과제	개선 핵심 사항
<ul style="list-style-type: none"> <li>• 무응답 현황 정리 및 분석</li> </ul>	<ul style="list-style-type: none"> <li>· 응답현황에 대한 표준화된 분류체계 개발. 무응답의 유형을 분류하여 자료값이 0(실적없음), 휴폐업, 거절, 연락 안 됨 등을 구분하여 현황을 파악하여 분석함</li> <li>· 표본 기관의 응답현황 모니터링 체계 구축</li> </ul>
<ul style="list-style-type: none"> <li>• 무응답 사업체에 대한 사후조사</li> </ul>	<ul style="list-style-type: none"> <li>· 무응답 사업체를 대상으로 부차표본으로 추출하여 재 조사를 통해서 무응답 보정 방안 마련</li> </ul>
<ul style="list-style-type: none"> <li>• 통계작성 공표단위 검토</li> </ul>	<ul style="list-style-type: none"> <li>· 현행 작성 통계의 공표단위를 살펴서 지나치게 세분된 것은 없는지 검토하여 필요한 경우에는 통합하는 방안 모색</li> </ul>



중기과제 (1년 이상)	
개선 과제	개선 핵심 사항
<ul style="list-style-type: none"> <li>• 표본추출틀 정비</li> </ul>	<ul style="list-style-type: none"> <li>· 표본추출틀의 기본 정보 보완(4.1절 내용 참고)</li> </ul>
<ul style="list-style-type: none"> <li>• 무응답에 대한 대체 방법 적용</li> </ul>	<ul style="list-style-type: none"> <li>· 응답거절 등의 단순 무응답의 경우는 비(ratio) 대체법(imputation)을 적용하여 대체하여 추정하는 방안 모색</li> </ul>
<ul style="list-style-type: none"> <li>• 조사데이터 신뢰성 검토</li> </ul>	<ul style="list-style-type: none"> <li>· 응답 기업체 중 일부를 부차표본으로 추출하여 측정오차를 분석함</li> </ul>
<ul style="list-style-type: none"> <li>• 특이치 검출 방법에 대한 기초 연구</li> </ul>	<ul style="list-style-type: none"> <li>· 특이치 검출법에 대한 기초 연구 진행</li> </ul>



장기과제 (3년 이상)	
개선 과제	개선 핵심 사항
<ul style="list-style-type: none"> <li>• 특이치 검출을 위한 시스템 구축</li> </ul>	<ul style="list-style-type: none"> <li>· Hidiroglou and Berthelot 방법, ADL 방법 등에 의한 특이치 검출 시스템을 구축하여 자동화된 특이치 검출 방안 마련</li> </ul>
<ul style="list-style-type: none"> <li>• 전수조사와 표본조사 병행</li> </ul>	<ul style="list-style-type: none"> <li>· 표본추출틀 정비를 기반으로 하며 표본설계를 통해 전수층과 표본층으로 구분하여 표본추출</li> </ul>
<ul style="list-style-type: none"> <li>• 영구난수법을 적용한 표본추출 방안 마련</li> </ul>	<ul style="list-style-type: none"> <li>· 정비된 표본추출틀을 토대로 영구난수법(PRN 방법)을 적용한 표본추출 방안 마련</li> <li>· 신설 기업체의 체계적인 표본 반영</li> <li>· 표본층에 속한 기업체의 응답부담 경감 가능</li> </ul>

## 참 고 문 헌

1. 국가과학기술위원회 (2012) 「2011년도 연구개발활동조사보고서」
2. 신동완, 김재광 (2006) 「과학기술연구개발조사 무응답처리」, 과학기술부
3. 유승훈, 문혜선 (2002) 과학기술연구개발활동조사의 개선방안, 기술혁신학회지, 제5권 제2호, pp. 228-244.
4. 이의규 (2010) 주기적 조사 자료의 내검 : 그래프 활용을 중심으로, 통계연구 15, 16-27.
5. 조성표, 박선영, 한기인, 노민선, 배한수, 김현아 (2009) 기업연구개발활동통계 개선방안에 관한 연구, 기술경영경제학회 2009년도 동계학술발표회 자료집: 313-332.
6. 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*.
7. 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.
8. Bethlehem, J. (2009) *Applied survey methods*, Wiley.
9. Hansen, M. H. and Hurwitz, W. N. (1946) The problem of non-response in sample surveys. *J. Am. Stat. Assoc.* 41:517-529.
10. Hidirolou, M. A. and Berthelot, J. M. (1986) Statistical edit and imputation for periodic surveys, *Survey Methodology* 12, 73-83.
11. Hulliger, B. (1999) Simple and robust estimators for sampling, *Proceedings of the survey research methods section, American Statistical Association*, 54-63.
12. Lavalee, P. and Hidirolou, M. A. (1988) On the Stratification of Skewed Populations. *Survey Methodology*, Vol. 14, No. 1, pp. 33-43.

13. Rivest, L.-P. (1999) Stratum jumpers: Can we Avoid Them?. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 64-72.
14. Rouseeuw and Zomeren (1990) Unmasking multivariate outliers and leverage points, *Journal of the American Statistical Association* 85, 633-639.
15. Srinath, K.P. and Carpenter, R.M. (1995). "Sampling Methods for Repeated Business Surveys," *Business Survey Methods*, Eds. Cox, B.G. et al., Wiley.
16. 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*.
17. 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*.

---

# 부 록

---



부록 1. 2011년도 「연구개발활동조사」의  
산업중분류 및 규모별 응답 기업체 현황

부록 2. 2011년도 「연구개발활동조사」의  
산업중분류 및 규모별 분석 결과

부록 3. 모집단 분석: 산업중분류 및 규모

부록 1. 산업중분류 및 기업체 규모별 응답 기업체 현황: 2011년도 조사

산업중 분류	4인 이하	5-9인	10-49인	50-99인	100-299인	300인 이상	총합
01	8	11	2	4	1	0	26
05	0	0	0	0	0	1	1
10	64	225	91	80	17	15	492
11	6	15	9	4	6	2	42
12	0	0	0	0	0	1	1
13	30	146	66	46	7	1	296
14	10	87	39	33	16	0	185
15	5	29	6	8	3	0	51
16	6	34	6	6	0	0	52
17	6	54	16	12	7	1	96
18	13	28	17	6	1	1	66
19	5	16	5	3	0	3	32
20	219	609	128	99	41	19	1,115
21	30	99	35	69	35	9	277
22	45	302	115	101	16	6	585
23	31	152	39	30	17	4	273
24	15	149	81	72	19	12	348
25	88	540	122	80	18	2	850
26	451	1,205	271	265	57	26	2,275
27	216	565	124	63	8	1	977
28	244	716	156	97	26	4	1,243
29	382	1,616	359	208	39	11	2,615
30	23	301	172	198	73	22	789
31	15	89	40	28	9	9	190
32	8	71	15	7	5	1	107
33	92	262	30	16	0	0	400
35	9	14	1	1	3	6	34
36	1	0	0	0	0	0	1
37	3	15	0	0	1	1	20
38	8	48	15	2	1	0	74
39	4	13	1	0	0	0	18
41	51	854	86	36	19	19	1,065
42	87	377	59	29	9	0	561

(계속)

산업중 분류	4인 이하	5-9인	10-49인	50-99인	100-299인	300인 이상	총합
45	1	3	3	1	2	0	10
46	110	456	89	37	7	5	704
47	9	23	3	4	3	0	42
49	0	3	0	0	0	2	5
51	0	0	0	0	0	1	1
52	0	5	0	4	2	0	11
55	0	1	0	0	0	1	2
56	0	2	0	1	1	1	5
58	213	892	192	123	21	6	1,447
59	2	17	6	3	0	0	28
60	0	1	1	0	2	1	5
61	6	19	4	4	1	3	37
62	125	357	51	27	7	4	571
63	32	91	29	15	1	2	170
64	0	1	0	0	0	0	1
66	0	0	0	1	0	0	1
68	2	3	1	0	0	0	6
69	0	2	1	1	1	0	5
70	131	155	21	5	0	0	312
71	9	39	11	7	1	2	69
72	83	463	171	122	38	12	889
73	47	87	7	4	1	0	146
74	2	8	1	4	2	2	19
75	0	13	7	3	2	5	30
85	3	21	4	6	3	0	37
86	0	1	0	0	0	0	1
90	4	2	0	0	0	0	6
91	0	2	0	0	0	1	3
95	2	11	4	3	4	0	24
96	1	3	0	2	0	0	6
총합	2,957	11,323	2,712	1,980	553	225	19,750

## 부록 2. 2011년도 「연구개발활동조사」의 산업중분류 및 규모별 분석 결과

<부표 2-1> 산업중분류 및 기업체 규모별 연구개발비 총액 현황: 2011년도 조사

산업대분류	산업중분류	10인 미만	10-29인	30-99인	총합
C. 제조업	10	124.4	526.5	233.9	884.8
	11	11.1	50.5	38.1	99.6
	12	-	-	-	-
	13	61.1	465.6	211.7	738.4
	14	37.2	184.1	157.5	378.8
	15	4.5	46.0	66.8	117.3
	16	26.9	91.0	18.4	136.3
	17	12.2	133.4	55.8	201.3
	18	9.8	56.8	39.6	106.2
	19	2.6	62.4	35.9	100.9
	20	474.4	2,370.5	1,015.6	3,860.5
	21	74.7	556.8	422.3	1,053.7
	22	54.9	889.3	557.3	1,501.5
	23	43.1	403.2	188.9	635.2
	24	18.1	372.5	336.4	727.0
	25	109.4	1,459.9	534.5	2,103.7
	26	1,041.4	6,457.1	3,266.6	10,765.1
	27	499.6	2,793.1	1,437.5	4,730.2
	28	406.3	2,515.8	1,205.9	4,128.1
	29	704.0	5,796.8	2,729.5	9,230.3
30	28.1	1,394.1	1,229.2	2,651.4	
31	28.0	372.1	222.0	622.1	
32	8.5	138.0	64.0	210.5	
33	128.9	712.6	146.9	988.4	
F. 건설업	41	86.9	3,349.5	851.6	4,288.0
	42	67.4	769.6	257.8	1,094.9
J. 출판, 영상, 방송통신 및 정보서비스	58	306.9	3,093.7	1,297.2	4,697.8
	59	2.6	47.2	48.9	98.7
	60		7.5	4.4	11.9
	61	12.6	52.2	47.1	111.8
	62	177.2	1,015.9	259.1	1,452.2
63	39.8	300.7	163.3	503.8	
M. 전문, 과학 및 기술 서비 스업	70	365.7	1,579.9	431.8	2,377.4
	71	12.5	94.5	250.7	357.6
	72	113.5	1,236.3	595.8	1,945.6
	73	69.4	279.0	41.1	389.5
	총합	5,163.3	39,674.1	18,462.8	63,300.2

<부표 2-2> 산업중분류 및 기업체 규모별 연구개발비 표본표준편차: 2011년 조사

산업대분류	산업중분류	10인 미만	10-29인	30-99인	총합
C. 제조업	10	2.836	2.895	2.919	2.889
	11	2.314	3.753	3.890	3.551
	12	-	-	-	-
	13	4.927	4.217	2.426	3.918
	14	5.940	1.765	3.801	3.035
	15	0.510	1.449	22.587	8.897
	16	8.597	3.249	1.980	4.094
	17	2.771	3.537	3.559	3.477
	18	0.573	3.562	1.517	2.662
	19	0.369	2.740	3.368	3.284
	20	3.336	4.713	9.304	5.572
	21	2.922	6.344	15.236	9.177
	22	1.097	5.750	4.410	5.255
	23	1.843	3.309	4.968	3.636
	24	1.051	2.529	4.202	3.244
	25	1.035	5.336	4.372	4.939
	26	2.250	6.908	19.116	9.522
	27	2.617	6.760	17.766	8.985
	28	2.181	4.287	9.307	5.298
	29	2.384	5.158	9.312	5.931
30	0.991	12.688	18.652	14.833	
31	2.202	3.981	5.926	4.569	
32	0.626	2.601	3.565	2.808	
33	1.642	3.894	5.907	3.792	
F. 건설업	41	2.917	4.972	25.468	8.979
	42	0.666	2.205	3.982	2.493
J. 출판, 영상, 방송통신 및 정보서비스	58	1.565	4.559	6.787	4.872
	59	1.513	2.039	12.227	6.324
	60		0.000	0.000	2.164
	61	1.650	2.422	6.799	4.433
	62	1.413	3.571	4.549	3.448
63	0.896	5.460	4.437	4.850	
M. 전문, 과학 및 기술 서비 스업	70	2.493	19.028	28.042	16.152
	71	0.841	3.377	32.377	15.927
	72	0.945	2.959	4.172	3.200
	73	1.105	3.692	4.895	3.310
	총합	2.337	5.902	12.429	7.171

<부표 2-3> 산업중분류 및 기업체 규모별 연구개발인력 총수: 2011년 조사

산업대분류	산업중분류	10인 미만	10-29인	30-99인	총합
C. 제조업	10	211	1,027	533	1,771
	11	23	82	59	164
	12				
	13	86	691	377	1,154
	14	40	497	347	884
	15	10	142	46	198
	16	17	122	38	177
	17	26	215	88	329
	18	29	111	92	232
	19	10	86	40	136
	20	729	3,605	1,372	5,706
	21	139	841	448	1,428
	22	128	1,260	695	2,083
	23	89	648	275	1,012
	24	40	623	459	1,122
	25	247	2,217	814	3,278
	26	1,683	8,749	3,900	14,332
	27	821	3,709	1,756	6,286
	28	782	3,814	1,512	6,108
	29	1,179	8,482	3,624	13,285
F. 건설업	30	54	1,376	1,297	2,727
	31	49	573	342	964
	32	18	276	101	395
	33	282	1,297	279	1,858
J. 출판, 영상, 방송통신 및 정보서비스	41	162	4,150	617	4,929
	42	254	1,826	575	2,655
	58	873	8,068	3,291	12,232
	59	8	102	49	159
	60		12	11	23
	61	26	112	80	218
M. 전문, 과학 및 기술 서비 스업	62	458	2,397	659	3,514
	63	124	675	380	1,179
	70	617	1,721	663	3,001
	71	36	209	173	418
총합	72	317	2,881	1,280	4,478
	73	182	609	92	883
	총합	9,749	63,205	26,364	99,318

<부표 2-4> 산업중분류 및 기업체 규모별 연구개발인력 표본표준편차: 2011년 조사

산업대분류	산업중분류	10인 미만	10-29인	30-99인	총합
C. 제조업	10	1.578	2.578	3.349	2.767
	11	3.125	3.270	2.698	3.126
	12	-	-	-	-
	13	1.737	3.020	2.423	2.851
	14	1.826	3.144	6.766	4.683
	15	0.707	3.277	3.983	3.471
	16	1.472	1.743	3.559	2.211
	17	2.338	2.176	2.989	2.424
	18	1.092	1.972	2.623	2.309
	19	1.225	2.754	2.000	3.011
	20	1.722	3.799	7.234	4.631
	21	1.752	6.671	7.999	6.879
	22	1.581	2.101	2.960	2.494
	23	1.586	2.068	4.334	2.838
	24	1.234	2.278	3.102	2.666
	25	1.500	2.267	3.890	2.757
	26	1.695	5.117	13.293	7.201
	27	1.920	3.915	9.639	5.716
	28	1.946	3.301	7.512	4.392
	29	1.670	2.899	7.011	4.227
	30	1.071	2.519	5.234	3.974
31	1.534	4.485	6.652	5.188	
32	1.035	2.122	3.262	2.551	
33	1.728	3.678	7.052	4.002	
F. 건설업	41	1.786	3.281	3.549	3.334
	42	1.424	2.480	8.999	4.128
J. 출판, 영상, 방송통신 및 정보서비스	58	1.653	6.463	13.997	8.463
	59	2.828	2.937	4.834	3.510
	60		0.000	0.000	0.707
	61	2.160	3.298	13.736	7.366
	62	1.502	4.305	12.682	5.823
63	2.091	5.387	8.204	6.279	
M. 전문, 과학 및 기술 서비스업	70	1.734	8.003	17.040	9.858
	71	1.658	2.661	13.624	7.394
	72	1.795	3.139	6.540	4.235
	73	1.872	4.827	8.572	4.821
	총합	1.781	4.245	9.235	5.494

### 부록 3. 산업중분류 및 기업체 규모별 모집단 현황

<부표 3-1> 산업중분류 및 기업체 규모별 모집단 현황

산업중분류	연구원, 교 원 창업기업	벤처기업	소기업	중기업	대기업	총합
01	0	6	10	15	6	37
02	0	1	3	1	0	5
03	0	3	0	1	0	4
05	0	0	1	0	1	2
07	0	0	5	0	0	5
10	0	182	542	188	75	987
11	0	15	28	12	14	69
12	0	0	1	1	2	4
13	0	93	354	132	25	604
14	0	33	321	103	15	472
15	0	19	92	23	6	140
16	0	12	42	7	3	64
17	0	27	110	49	11	197
18	0	30	118	23	2	173
19	0	6	24	4	8	42
20	1	475	997	218	169	1,860
21	1	139	145	115	81	481
22	0	245	570	229	46	1,090
23	0	96	302	65	52	515
24	0	107	284	148	71	610
25	4	402	1,055	236	52	1,749
26	3	1,675	1,986	505	264	4,433
27	1	686	821	108	33	1,649
28	1	751	1,324	242	110	2,428
29	3	1,441	2,773	464	158	4,839
30	0	239	634	449	208	1,530
31	1	108	173	76	44	402
32	0	31	143	26	11	211
33	0	145	480	52	9	686
35	0	7	12	13	14	46
36	0	1	1	0	1	3
37	0	4	7	15	2	28
38	0	13	19	66	7	105
39	0	17	11	10	0	38
41	0	61	1,303	142	91	1,597
42	1	122	686	79	17	905

(계속)

산업중분류	연구원, 교 원 창업기업	벤처기업	소기업	중기업	대기업	총합
45	0	2	6	11	0	19
46	0	155	432	542	19	1,148
47	0	26	47	59	2	134
49	0	1	4	3	10	18
50	0	0	1	0	1	2
51	0	0	0	0	2	2
52	0	0	8	5	5	18
55	0	0	0	0	1	1
56	0	1	1	7	1	10
58	2	1,221	1,178	402	78	2,881
59	0	20	25	7	1	53
60	0	2	3	2	8	15
61	0	29	27	12	11	79
62	2	551	680	147	34	1,414
63	0	113	123	54	9	299
64	0	0	0	1	0	1
66	0	2	2	1	1	6
68	0	1	0	0	2	3
69	0	0	3	3	2	8
70	4	178	129	15	13	339
71	0	30	78	20	4	132
72	2	263	628	449	59	1,401
73	1	61	137	33	6	238
74	0	7	7	10	3	27
75	0	16	23	9	5	53
85	0	12	12	18	2	44
86	0	2	13	8	1	24
90	0	0	2	4	0	6
91	0	0	2	1	0	3
94	0	0	0	1	0	1
95	0	0	0	5	2	7
96	0	0	3	5	0	8
총합	27	9,885	18,951	5,651	1,890	36,404