

연구노트

사업체패널조사의 조사설계

Survey Design of the Workplace Panel Survey in Korea

이기재*, 김혜원**, 김수진***, 김기민****, 이용희*****

Kee Jae Lee · Hye Won Kim · Sue Jin Kim · Ki Min Kim · Yong Hee Lee

사업체패널조사는 상용근로자 30인 이상의 사업체를 대표하는 패널을 구성하여 전반적인 경영환경 및 인적자원 관리체계, 노사관계의 현황 및 임금교섭 과정 등에 대한 정보를 추적·조사하는 종단면 조사이다. 이 조사는 우리나라의 고용구조 및 노동수요를 체계적으로 파악하고, 전반적인 인적자원 관리체계를 평가하여 바람직한 노동시장정책을 수립하기 위한 기초자료로 활용하는 것을 목적으로 한다. 조사에서는 인사담당자 응답용, 노사관계 업무담당자 응답용, 근로자대표 응답용으로 구분된 세 종류의 설문지가 사용되었다.

사업체패널조사의 표본사업체는 지역, 업종 구분 및 사업장 규모 등을 층화변수로 하는 층화임의추출법에 따라 추출되었다. 사업체패널조사의 가중치는 설계 가중치, 무응답 조정, 사후층화 조정의 과정을 거쳐 작성되었으며, 조사 실행과정은 조사 데이터의 질을 높이고자 컴퓨터보조개별면접(CAPI) 방법을 적용하여 진행되었다.

본 연구에서는 사업체패널조사에 대한 구조와 주요 설문내용, 표본설계, 실사과정, 응답률 현황, 무응답 대체방법, 가중치 작성과정, 조사결과에 대한 설계효과 분석 등에 대해서 논의하였다.

주제어 : 패널조사, 층화임의추출법, 가중치, 무응답 대체, 컴퓨터보조개별면접, 설계효과

* 교신저자(corresponding author): 한국방송통신대학교 정보통계학과 교수 이기재,
E-mail: kjlee@knou.ac.kr

** 한국노동연구원 연구위원.

*** 한국노동연구원 전 연구원.

**** 한국노동연구원 연구원.

***** 서울시립대학교 통계학과 조교수.

Workplace Panel Survey(WPS) is the representative panel survey of workplace in Korea. WPS was newly sampled in 2005 and is to be used for the subsequent biennial survey. The main survey is divided into a questionnaire for human resources(HR) manager, a questionnaire for labor relations manager and a questionnaire for representatives of unions.

The population of WPS 2005 included workplaces across the country with 30 or more employees. The WPS 2005 was composed of 1,905 workplaces including 290 workplaces in the public sector. The sample was selected by the stratified random sampling. Weighting process for the survey data was introduced to compensate for differential sampling and non-response rates.

Personal interviews were conducted using the Computer Assisted Personal Interviewing(CAPI) system during visits by interviewers, along with survey via mail and e-mail concerning employment and financial issues. The CPAI system introduced for the WPS 2005 can be used for automatical detection for errors and inconsistencies which may occur during the survey process. The CAPI system played an important part in enhancing the reliability of the survey data.

Key words : panel survey, stratified random sampling, weight, imputation, CAPI, design effect

I . 서론

사업체패널조사(WPS : Workplace Panel Survey)는 일정 규모 이상의 사업체를 대표하는 패널을 구성하여 전반적인 경영환경 및 인적자원 관리체계, 노사관계의 현황 및 임금교섭 과정 등에 대한 정보를 추적·조사하는 사업체 대상의 종단면 조사이다. 이 조사는 우리나라의 고용구조 및 노동수요를 체계적으로 파악하고, 전반적인 인적자원 관리체계를 평가하여 바람직한 노동시장 정책을 수립하기 위한 기초자료로 활용하는 것을 목적으로 한다. 사업체패널 조사의 세부 목적을 정리하면 다음과 같다.

첫째, 사업체 단위의 고용구조 및 노동수요를 파악하는 것이다. 특히 경제 위기 이후 급변한 기업의 노동수요 구조 및 변화를 읽고, 임시·일용직과 사내 하청 등 비정규직 현황을 파악하여 기업의 고용 유연화 전략과 그 특성을 알아 보고자 한다.

둘째, 인적자원관리 및 노사관계의 흐름과 그 변화를 파악하는 것이다. 이를 통해 임금체계, 평가 및 보상 등 전반적인 인적자원 관리구조 및 특성과 한국형 노사관계의 성격, 특징 그리고, 구조에 관해 실증적으로 연구하고자 한다.

셋째, 기업의 인적자원개발 및 투자실태를 분석한다. 즉, 교육훈련과 기업 내 숙련인력의 형성 등 인적자원개발과 투자의 실태 및 변화를 파악할 필요가 있다.

마지막으로, 해외 사업체 조사와의 협력관계를 형성함으로써 국제적으로 비교 가능한 사업체 관련 자료를 구축하고자 한다. 이는 국가 간 비교연구뿐 아니라 학제 간 연구의 폭을 증대시켜 다양한 분석과 연구를 가능하게 할 것이다.

이러한 목적을 위하여 사업체패널조사의 설문은 인사담당자 응답용, 노사 관계 업무담당자(또는 노사협의회 업무담당자) 응답용, 근로자대표(또는 노사협의회 근로자대표) 응답용 등으로 구분하여 작성되었다(〈그림 1〉 참조).

우리나라의 사업체패널과 유사한 사례로는 캐나다의 WES(Workplace and Employee Survey)와 영국의 WERS(Workplace Employment Relations Survey) 등을 들 수 있다. 두 조사의 표본설계는 Patak et al.(1998)과 Chaplin et al.(2005)에 제시되어 있다.

한국노동연구원은 2002년부터 2004년까지 3차례에 걸쳐 사업체패널조사를 실시하였다. 세 차례에 걸친 사업체패널조사는 우리나라 사업체의 인적자원관리와 노동수요 및 노사관계 등의 다양한 이슈에 대한 국내 유일의 대규모 데이터를 제공하였으나, 체계적인 설문 및 표본설계 과정이 부족하였다는 지적과 더불어 조사 데이터의 불일치 및 논리오류 등의 문제점이 제기되었다. 이에 따라 한국노동연구원은 좀 더 양질의 사업체 관련 데이터를 생산하고 제공하기 위하여 지난 세 차례에 걸친 사업체패널조사를 분석하여 문제점과 시사점을 파악하고, 외국의 사업체패널조사 사례를 연구하였다. 한국노동연구원은 다양

한 분야의 전문가로 구성된 연구그룹을 구성하여 1년여 동안 사업체패널조사의 설문지를 대폭 개편하고, 새로운 표본설계와 조사방법(컴퓨터보조면접조사)을 도입하는 준비과정을 거쳐 2006년에는 새로운 2,000여개 사업장들로 구성된 패널을 구축하여 조사를 실시하였다.

본 논문은 사업체패널조사의 조사설계와 실사과정을 소개함으로써 다양한 분야의 연구자들이 조사결과를 올바르게 효율적으로 활용할 수 있도록 하는데 그 목적이 있다. 이러한 목적을 위해, 본 논문에서는 사업체패널조사에 대한 구조와 주요 설문내용, 표본설계, 실사과정, 응답률 현황, 무응답 대체방법, 가중치 작성과정, 조사결과에 대한 설계효과 분석 등에 대해서 살펴본다.



〈그림 1〉 사업체패널조사의 설문구조

II. 표본설계

1. 조사모집단

사업체패널조사의 조사단위는 원칙적으로 사업체이지만, 금융 및 보험업, 통신업, 전기, 수도 및 가스업 등은 개별 사업장이 아닌 본사를 조사대상으로 하였다. 이는 이들 산업에서는 인사 및 노사관계에 대한 관리가 본사 차원에서 이루어지고 있다는 점을 고려한 것이다.

사업체패널조사의 추출틀은 표본설계 당시의 최신 자료인 2004년 12월 말 기준의 사업체기초통계조사(통계청) 자료를 이용하였다. 조사대상은 크게 일반사업장과 공공사업장의 두 부문으로 구성된다. 먼저 일반사업장 부문은 농·임업, 어업, 광업 등의 산업을 제외한 상용근로자 기준으로 30인 이상의 사업장들이 포함되었다. 다만, 종교단체(9191), 아파트 관리사무소(70211), 주점업(5523), 미용, 욕탕 및 유사 서비스업(931), 회원단체(91) 등은 다른 일반적인 사업장과 그 성격이 크게 다르기 때문에 조사대상에서 제외하였다.¹⁾ 한편, 사업체기초통계조사 상의 조직형태 구분은 개인사업체, 회사법인, 회사 이외의 법인, 비법인 단체 등으로 구분된다.²⁾ 이들 중 비법인 단체로 구분된 147개 사업장은 조사대상에서 제외하였다. 이들 사업장은 대부분 사업서비스업이나 사회서비스업에 속한다. 조사대상에서 제외된 사업장은 전체 조사대상

1) 종교단체(9191), 아파트 관리사무소(70211), 주점업(5523), 미용, 욕탕 및 유사 서비스업(931), 회원단체(91) 등에서 괄호 안의 값은 한국표준산업분류 번호를 의미한다. 회원단체(91)에는 산업 및 전문가 단체(911), 노동조합(912), 기타회원단체(919) 등이 있다.

2) 개인사업체는 순수하게 개인이 사업을 경영하는 경우로 동업(공동경영)의 경우도 포함한다. 회사법인은 상법의 규정에 의하여 설립된 회사를 말하며, 주식회사, 유한회사, 합자회사, 합명회사가 있으며, 공기업이라도 주식회사의 형태를 취하고 있으면 해당된다. 회사 이외 법인은 민법 또는 특별법에 의하여 설립된 회사 이외의 법인을 말하며 재단법인, 학교법인, 의료법인, 사회복지법인, 종교법인 등이 있다. 비법인 단체에는 법인격이 없는 각종 협회, 조합, 동창회, 후원회, 문화단체, 노동단체, 종친회, 교회 등이 있다.

에서 사업장 기준으로 2.6%(1,005개소)이다.

공공사업장 부문의 조사모집단은 기획예산처가 정한 공공기관 구분기준에 따라 결정되었다. 기획예산처는 원칙적으로 정부산하기관관리기본법에서 규정한 공공기관 구분기준(법 제3조)을 최대한 적용하되 예외범위를 축소하여 총 314개 공공기관을 선정하였다. 여기에는 정부가 최대 지분을 소유하는 출자기관, 법률에 정부 출연근거가 있는 출연기관, 정부지원이 총수입의 50% 이상인 보조위탁기관, 출자·출연·보조위탁기관이 최대 지분을 소유한 자회사, 출연기관의 부설기관으로 정부출연을 받은 재출연기관, 기타 정부산하기관 운영위원회가 정하는 경영혁신대상기관 등이 포함된다.

공공 부문의 최종 조사모집단은 기획예산처가 선정한 314개 전체 공공기관과 51개의 지방 공기업을 합쳐 전체 365개 공공기관으로 하였다. 공공 부문의 조사대상은 일반사업장 부문의 조사대상과는 달리 상용근로자 기준으로 20인 이상의 공공사업장으로 하였다.

2. 층화

사업체패널조사는 사업장을 조사단위로 하며 다수의 조사항목을 조사하는 다목적 조사이다. 이 조사를 위한 표본설계는 통계적으로 신뢰성 높은 산업별, 사업장 규모별 그리고, 지역별 통계생산을 통해서 이들 간의 비교가 가능하도록 하는 데에 주안점을 두었다.

사업체패널조사에서 층화는 지역, 업종 및 사업장 규모 등을 이용하였다. 지역 구분은 서울권, 경기·인천권, 강원·충청권, 전라·제주권, 영남권 등 전국을 5개 권역으로 구분하였다. 사업체의 업종은 산업의 특성에 따라서 12개 업종으로 구분하였다(〈표 1〉 참조).

〈표 1〉 표본설계에서 사용된 업종 구분 현황

업종 구분		해당 산업 중분류 또는 대분류 ³⁾
제조업	경공업	15~22, 36
	화학공업	23~26, 37
	금속, 자동차, 운송장비	27~29, 34, 35
	전기, 전자, 정밀공업	30~33
건설업		F. 건설업
서비스업	개인서비스업	G. 도매 및 소매업 H. 숙박 및 음식점업
		I. 운수업 J. 통신업
	유통서비스업	K. 금융 및 보험업
		L. 부동산업 및 임대업 M. 사업서비스업
	사업서비스업	O. 교육서비스업 P. 보건 및 사회복지산업 Q. 오락, 문화 및 운동 R. 기타, 개인서비스업
	사회서비스업	
기타	E. 전기, 가스 및 수도사업	

사업장 규모 구분은 사업장의 상용근로자⁴⁾ 수를 기준으로 30~99인, 100~299인, 300~499인, 500인 이상 등으로 구분하였다. 부분적으로 층 내의 모집단 사업체가 1개인 경우에는 동일 지역 및 업종 구분에서 300~499인 규모와 500인 이상 규모를 하나의 층으로 묶어서 층화하였다. 표본설계에서 사용된 조사모집단 현황은 〈표 2〉와 같다.

3) 〈표 1〉에 있는 두 자리 숫자는 한국표준산업분류 상의 산업중분류 번호이다.

4) 사업체에 상근 고용된 자로 고용주와 1년 이상의 고용계약을 맺거나, 기간을 정하지 않고 고용된 자로 일반적으로 정규직원이라고 하는 자를 말한다.

〈표 2〉 업종 구분 및 사업장 규모별 모집단 사업장 수 현황

(단위 : 개소)

산업분류		30~99인	100~299인	300~499인	500인 이상	계
경공업		3,835	779	112	62	4,788
화학공업		2,441	543	84	65	3,133
금속, 자동차, 운송		4,663	883	101	116	5,763
전기, 전자, 정밀공업		2,351	644	89	115	3,199
건설업		1,455	201	19	21	1,696
개인서비스업		2,800	507	59	45	3,411
유통서비스업	운수업	1,976	1,019	80	46	3,121
	통신업	89	23	12	10	134
사업서비스업	금융 및 보험업	192	51	34	44	321
	기타	3,453	886	173	134	4,646
사회서비스업		4,474	733	97	152	5,456
전기, 가스 및 수도사업		27	12	0	0	39
계		27,756	6,281	860	810	35,707

주 : 통신업, 금융 및 보험업과 전기 및 가스업은 지점 및 지사를 제외한 본사 기준 사업장 수

3. 표본크기 및 표본추출

사업체패널조사의 표본크기는 모비율 추정에 대한 오차의 한계를 95% 신뢰수준에서 전국 통계에 대해서 3% 내외, 사업체 수가 적은 일부 업종을 제외한 각 업종 구분에서 8~12% 수준을 만족할 수 있도록 결정되었다. 이와 같은 원칙에 따라서 결정된 조사성공 목표 사업체 수(공공 부문 사업체 제외)는 1,500개소이다⁵⁾.

우선 각 업종 구분별 통계작성을 위해서 통신업, 금융 및 보험업, 전기, 가스 및 수도업에 대한 표본추출률은 상대적으로 높게 정하였다(〈표 3〉 참조). 나머지 업종별 조사성공 목표 사업체 수는 각 업종별 통계작성이 가능하도록 우선 100~120개를 배분하고, 나머지는 각 업종별 사업체 수에 비례배분하였다. 이러한 표본배분 방식은 각 업종 구분별 통계작성을 통해서 업종별 비교가 중요하다는 점을 고려한 것이다. 이와 같은 방법으로 모비율 추정의 허용오차

5) 실제 표본추출은 각 층별로 배정된 조사성공 사업체 수를 만족할 수 있도록 전체 약 3,916개 표본 사업장을 추출하였다. 이 값은 사업체 규모 구분에 따라서 조사 성공률을 40~50%로 가정하여 구한 것이다.

의 한계는 대부분의 업종 구분에서 8~12%로 관리된다.

업종 구분 내의 사업체 규모별 표본배분 방법은 비례배분법, 제공근 배분법, Kish의 절충배분법(Kish 1987) 등을 검토하여 절충배분법을 사용하였다. 상대적으로 근로자 300인 이상을 고용하고 있는 사업체에 대한 추출률을 높여 사업체 규모별 통계의 안정적인 작성이 가능하도록 하였다. 업종×규모 구분에서 권역별 표본배분은 비례배분법을 적용하였다. <표 3>은 사업체패널조사의 표본 사업장 수 및 표본추출률 현황이다.

2005 사업체패널조사의 공공부문 표본은 기획예산처에서 선정한 314개 전체 공공기관 가운데 예비조사 표본 6개를 제외한 308개 공공기관과 51개의 지방 공기업을 대상으로 전수추출하였다.

<표 3> 표본 사업장 수 및 추출률 (공공부문 제외)

(단위 : 개소, %)

산업분류		30-99인	100-299인	300-499인	500인 이상	계
경공업		207 (5.4)	119 (15.3)	95 (84.8)	59 (95.2)	480 (10.0)
화학공업		140 (5.7)	98 (18.0)	74 (88.1)	56 (86.2)	368 (11.7)
금속, 자동차, 운송		214 (4.6)	116 (13.1)	75 (74.3)	99 (85.3)	504 (8.7)
전기, 전자, 정밀공업		147 (6.3)	100 (15.5)	80 (89.9)	98 (85.2)	425 (13.3)
건설업		170 (11.7)	97 (48.3)	13 (68.4)	15 (71.4)	295 (17.4)
개인서비스업		181 (6.5)	94 (18.5)	57 (96.6)	41 (91.1)	373 (10.9)
유통서비스업	운수업	106 (5.4)	85 (8.3)	49 (61.3)	33 (71.7)	273 (8.7)
	통신업	35 (39.3)	23 (100.0)	9 (75.0)	10 (100.0)	77 (57.5)
사업서비스업	금융 및 보험업	57 (29.7)	23 (45.1)	19 (55.9)	26 (59.1)	125 (38.9)
	기타	200 (5.8)	102 (11.5)	93 (53.8)	111 (82.8)	506 (10.9)
사회서비스업		182 (4.1)	94 (12.8)	79 (81.4)	112 (73.7)	467 (8.6)
전기, 가스 및 수도사업		13 (48.1)	10 (83.3)	-	-	23 (59.0)
계		1,652 (6.0)	961 (15.3)	643 (74.8)	660 (81.5)	3,916 (11.0)

주 : 통신업, 금융 및 보험업, 전기 및 가스업은 지점 및 지사를 제외한 본사 기준의 사업장 수

Ⅲ. 실사 및 응답현황

1. 실사

사업체패널조사의 응답자는 각 사업장의 인사업무 담당자와 노사관계 업무 담당자 및 근로자대표로 구성된다. 면접원이 응답자를 직접 방문하여 노트북 컴퓨터를 이용한 면접방식인 CAPI(Computer Assisted Personal Interview)를 통해서 조사가 진행되었다. 조사기간은 2006년 7월부터 11월까지 약 5개월 간으로 사전 접촉(contact), 우편물 발송, 방문 면접조사 등의 과정을 거쳤다.

사전 접촉은 훈련받은 1차 면접원이 표본사업장에 전화를 걸어 사업장명, 업종, 주소 등 사업장의 기본 정보를 확인한 후 실사에 사용될 우편물 및 조사표를 수령하고 면접조사에 응답할 담당자를 찾는 단계이다. 이 때 조사기준 시점인 2005년 12월 말을 기준으로 조사 부적격 대상(III장 2절 참조)은 조사대상에서 제외하였다.

다음 단계로 사전 접촉을 통해 확인된 담당자에게 사업체패널조사를 소개하는 안내문(brochure), 조사협조를 요청하는 공문, 재무현황 및 근로자현황 조사표를 발송하였다. 재무현황 및 근로자현황 조사표의 경우 방문 면접이 진행되기 전에 응답을 완료하여 수거하는 것을 원칙으로 하였다. 재무현황 및 근로자현황 조사표는 응답자가 직접 기입하는 방식으로 진행되며, 팩스로 발송하여 수거하거나, 우편물로 발송한 조사표를 면접원이 직접 수거하기도 하였다.

사전 접촉을 통해 응답자가 선정되고 재무현황 및 근로자현황 조사표 작성이 완료되면 면접원이 해당 사업장을 방문하여 면접조사를 진행하였다. 면접조사는 네덜란드 통계청이 개발한 Blaise 시스템을 활용한 CAPI를 사용하였고, 이 때 CAPI 면접방식은 면접원이 조사대상 사업장을 직접 방문하여 노트북 컴퓨터에 설문을 응답하도록 하는 개별 면접법이다.

2. 응답 현황

사업체패널조사는 359개 공공부문 사업장과 3,916개 일반사업장을 합쳐 총

4,275개 표본 사업장을 대상으로 진행하였다. 사업장 중 723개 사업장이 조사 부적격으로 조사대상에서 제외되었으며, 나머지 유효표본 3,552개 중 1,905개 사업장이 조사에 응답하였다. 나머지 1,647개 사업장에서는 응답거절 등의 사유로 조사에 성공하지 못하였다(〈표 4〉 참조). 일반 사업장의 비가중 응답률은 유효표본 3,224개 사업장 중 1,615개 사업장이 조사에 응답하여 50.1%이고, 설계가중치를 고려한 가중 응답률은 46.3%이다. 공공부문의 경우에는 31개 사업장이 조사대상에서 제외된 유효표본 328개 사업장 중 290개 사업장에서 응답하여 88.4%의 응답률을 나타냈다.

〈표 4〉 2005 사업체패널조사 표본 및 조사완료 현황

		전 체		공공부문 제외		공공부문		
		사업장 수	비율 (%)	사업장 수	비율 (%)	사업장 수	비율 (%)	
조사대상에서 제외	30인 미만 사업장*	393	9.2	368	9.4	25	7.0	
	해당 안 됨 (2006년 설립 등)	23	0.5	20	0.5	3	0.8	
	폐업/폐쇄/휴업/합병	63	1.5	62	1.6	1	0.3	
	결번	64	1.5	64	1.6			
	잘못된 번호	98	2.3	97	2.5	1	0.3	
	지방 공기업	30	0.7	30	0.8			
	리스트 중복	14	0.3	14	0.4			
	사업장/사업장명 다름	31	0.7	30	0.8	1	0.3	
	기타사항	7	0.2	7	0.2			
	소 계	723	16.9	692	17.7	31	8.6	
조사 실패	거절	1,283	30.0	1,245	31.8	38	10.6	
	연락 안 됨	연락처 불명	50	1.2	50	1.3		
		담당자 부재	45	1.1	45	1.1		
	기타 사항	내부결재 중	9	0.2	9	0.2		
		보류	192	4.5	192	4.9		
		진행중단	68	1.6	68	1.7		
	소 계	1,647	38.5	1,609	41.1	38	10.6	
조사 성공	1,905	44.6	1,615	41.2	290	80.8		
전 체	4,275	100.0	3,916	100.0	359	100.0		

* 공공부문의 경우 20인 미만 사업장.

사업체패널조사의 사업장 규모별 조사 부적격률과 응답률은 <표 5>에 제시되어 있다. 이 가운데 공공부문을 제외한 일반 사업장의 조사 부적격률은 17.7%이며, 특히 100인 미만 소규모 사업장에서 조사 부적격률이 높게 나타났다. 이는 소규모 사업장의 경우 표본추출 당시와 비교하여 조사 시점에 사업장 규모가 축소되어 30인 미만으로 분류되거나, 폐업·휴업 등의 사유가 많이 발생하기 때문이다. 공공부문은 일반 사업장에 비하여 사업장 특성 등에 있어서 변화가 적고 전화번호 등의 정확한 접촉을 위한 정보수집이 용이하기 때문에 조사 부적격률이 8.6%로 낮다. 조사 부적격률과 응답률을 계산할 때 설계가중치를 고려한 경우를 살펴보면 조사 부적격률은 다소 증가하고, 응답률은 다소 떨어지는 경향을 나타냈다. 이는 상대적으로 표본추출률이 낮은 업종이나 규모 구분에서 조사 부적격률은 높게 나타나고, 응답률은 다소 낮게 나타난 결과이다.

<표 5> 2005 사업체패널조사의 규모별 조사 부적격률 및 응답률

	전체	공공부문 제외					공공 부문	
		소계	30-99인	100-299인	300-499인	500인-		
최초 표본 (A)	4,275	3,916	1,652	961	643	660	359	
조사대상 제외 ⁶⁾ (B)	723	692	444	103	71	74	31	
조사 실패 (C)	1,647	1,609	586	450	248	325	38	
조사 성공 (D)	1,905	1,615	622	408	324	261	290	
조사 부적격률	비가중 ⁷⁾	16.9	17.7	26.9	10.7	11.0	11.2	8.6
	가중	24.4	24.5	28.2	11.4	12.7	11.4	8.6
응답률	비가중	53.6	50.1	51.5	47.6	56.6	44.5	88.4
	가중	46.9	46.3	47.0	43.6	53.1	40.8	88.4

주 : 조사 부적격률 = $\frac{B}{A} \times 100(\%)$, 응답률 = $\frac{D}{C+D} \times 100(\%)$

6) <표 5>와 <표 6>에서 조사대상 제외, 조사 실패 등의 구분은 <표 3>에 제시된 기준에 따라 구분된 것이다.

7) <표 5>와 <표 6>에서 조사 부적격률과 응답률 계산에서 '비가중'은 단순 비율을 구한 것이고, '가중'은 설계가중치를 고려하여 구한 결과이다.

〈표 6〉 2005 사업체패널조사 공공부문 제외 사업장의 산업별 조사 부적격률 및 응답률

	전체	제조업				전기 가스 수도업	건설업	개인 서비스업	운수업	통신업	금융 보험업	사업 서비스업 기타	사회 서비스업	
		경공업	화학 공업	금속 자동차 운송	전기 전자 정밀									
최초 표본(A)	3,916	480	368	504	425	23	295	373	273	77	125	506	467	
조사대상 제외(B)	692	100	44	69	73	0	76	69	49	22	14	123	53	
조사 실패(C)	1,609	195	173	226	191	3	112	152	85	22	53	198	199	
조사 성공(D)	1,615	185	151	209	161	20	107	152	139	33	58	185	215	
조사 부적격률	비가중	17.7	20.8	12.0	13.7	17.2	0	25.8	18.5	17.9	28.6	11.2	24.3	11.3
	가중	24.5	31.3	17.6	22.9	26.3	0	30.5	27.3	19.8	31.0	13.5	30.2	17.6
응답률	비가중	50.1	48.7	46.6	48.0	45.7	87.0	48.9	50.0	62.1	60.0	52.3	48.3	51.9
	가중	46.3	44.0	45.6	37.2	43.4	90.3	49.0	47.3	58.6	64.0	50.9	43.5	50.2

주 : 조사 부적격률 = $\frac{B}{A} \times 100(\%)$, 응답률 = $\frac{D}{C+D} \times 100(\%)$

공공부문을 제외한 사업장 가운데 조사 부적격률은 건설업이 25.8%로 가장 높았고, 전기·가스·수도업은 0%로 가장 낮게 나타났다. 이는 업종의 특성상 건설업이 사업장 상황의 변동이 가장 크고, 전기·가스·수도업은 상대적으로 적기 때문이다. 응답률의 경우에는 전기·가스·수도업이 87.0%로 가장 높았고, 운수업과 통신업이 각각 62.1%와 60.0%, 나머지 업종에서는 46~52% 정도로 나타났다. 조사 부적격률과 응답률을 계산할 때 설계가중치를 고려한 경우를 살펴보면 앞선 〈표 5〉의 결과와 마찬가지로 조사 부적격률은 다소 증가하고, 응답률은 다소 떨어지는 경향을 보이고 있다.

3. 데이터 편집

CAPI 시스템을 이용한 사업체패널조사는 실사가 진행됨에 따라 실사 현장에서 데이터의 논리 오류나 불일치성에 대하여 CAPI시스템에 입력된 편집규칙에 따른 검토가 가능하다. 한국노동연구원의 'Data Cleaning 가이드'를 기초

로 근로자 현황 조사와 인사담당자, 노무담당자 및 근로자대표 설문지의 오류를 검증할 수 있는 CAPI 시스템을 구축하였다. 우선, 각 설문마다의 오류 및 불일치성을 검증한 후 그에 대해 다른 보충 데이터나 혹은 재점검을 통해 오류를 수정하거나 결측치로 처리하였다. 또한 해당 표본 사업장에서 얻어진 설문을 서로 연결하여 설문 간의 오류 및 불일치성을 검증하여 오류를 수정하거나 결측치로 처리하였다. 이외에도 해당 표본 사업장의 산업분류와 ‘재무 현황 조사(FPQ)’ 결과의 정확성을 높이고자 고용보험 DB 정보, 한국신용평가정보 데이터 및 대한상공회의소 자료 등을 비교·검토하여 필요한 경우 재조사하거나 조사결과를 수정하였다.

4. 무응답 대체

사업체패널조사에서는 전체 4,275개의 표본 사업장 가운데 723개 사업장은 휴·폐업, 사업장 규모 과소 등의 이유로 조사대상 사업장에서 제외되었고, 1,647개의 사업장은 거절, 연락 두절 등의 사유로 조사를 성공하지 못하여 단위무응답(개체 무응답, unit nonresponse)이 발생하였다. 또한 조사에 협조한 1,905개의 사업장 중에서도 일부 사업장은 조사항목 중 일부를 응답하지 않거나 거부하는 항목무응답(item nonresponse)이 발생하였다. 전체 조사항목 중 약 190개의 항목에서 항목무응답이 발생하였지만 실제로 사업장에서 민감하게 생각하는 일부 설문(예: 영업이익, 임금인상을 및 결정방식 등)을 제외하면 무응답 비율은 5% 미만으로 낮은 편이다.

무응답 대체(imputation)는 항목무응답이 발생했을 때 실제 관측되지 않은 응답의 값에 대한 예측값을 구하여 대체함으로써 조사자료의 활용도를 높이는 방법을 말한다. 자료를 수집하여 제공하는 통계 생산자의 입장에서는 데이터 이용자가 무응답 대체과정을 쉽게 이해할 수 있는 간단하고 명확한 무응답 대체방법을 선호한다. 사업체패널조사에서는 대부분의 조사항목에서 무응답률이 대단히 낮아서 외부 정보 또는, 그 동안의 경험을 감안하여 조사에서 발생한 항목무응답을 무시가능 무응답(ignorable nonresponse)으로 간주하여 대체하였다.

무응답의 대체는 전체 표본을 여러 개의 대체군(imputation cell)으로 분할

한 뒤, 무응답이 발생한 해당 단위가 속한 대체군 내에서만 확률적 또는 통계적 모형의 추정 또는 예측을 이용하는 대체법을 이용하여 조건적으로 이루어졌다. 사업체패널조사에서 대체군을 구성하기 위해 사용한 변수는 사업장의 업종 구분과 사업장 규모이다. 사업장의 업종이 동일하고, 규모가 비슷할수록 사업장의 유사성이 증가할 것이다.

사업체패널조사에서 무응답의 대체는 조사항목의 특성과 제약조건을 고려하여 여러 가지 다른 방법들이 사용되었다. 예를 들어 재무 현황 관련 변수들(예를 들어 자산, 부채, 자본, 매출액, 총이익, 순이익, 각종 비용에 대한 항목 등)은 일반적으로 서로 확률적으로 관련되어 있으며, 어떤 변수들은 결정론적인 함수관계를 갖고 있다. 따라서 여러 개의 변수를 동시에 고려한 다변량 평균 대체법 등을 고려할 수 있지만 현실적으로 적용하기에는 어려움이 많다. 따라서 사업체패널조사에서 재무 현황 관련 변수에 대한 무응답 대체는 상황에 따라 핫덱대체법(hot deck imputation), 보조변수를 이용한 비추정법(ratio estimation), 모형을 이용한 확률적 대체법 등이 사용되었다. 예를 들어 사업장 중에서 모든 재무 관련 변수들에 무응답을 한 경우에는 핫덱대체법을 적용하였다. 핫덱대체법을 적용하여, 재무 변수들에 대한 무응답을 대체한 이후에 항목무응답 비율이 높지 않은(예로, 5% 이내) 조사항목에 대해서는 더 이상 무응답을 대체하지 않았다. 핫덱대체법을 적용한 후에도 항목무응답의 비율이 높은 조사항목에 대해서는 관련된 변수를 보조변수로 하는 비추정법을 이용하는 무응답 대체법에 따라 대체하였다. 이때 변수간의 함수적 관계와 제약식을 고려하였다.

IV. 가중치 작성

1. 횡단면 가중치

사업체패널조사의 횡단면 가중치는 설계가중치, 무응답 조정, 사후층화 조정의 과정을 거쳐 작성되었다. 먼저 표본 사업장의 설계가중치(design

weight)는 II장 3절에서 기술한 증화임의추출법에 따른 표본 사업장 추출확률의 역수로 정의된다.

무응답 조정은 사업장 특성에 따라서 다르게 나타날 수 있는 무응답 성향을 조정함으로써 추정 결과의 편향(차우침)을 줄이기 위함이다. 사업체패널조사에서 사업장 특성에 따른 응답률의 차이는 일부 산업을 제외하면 크지 않았다. 부분적으로 공공부문 사업장과 전기·가스·수도업의 경우에 다른 산업에 비해서 응답률이 높게 나타났지만, 나머지 업종에서는 응답률의 차이가 크지 않았다. 지역 구분과 사업장 규모 구분에 따른 응답률의 차이도 크지 않았다.

사업체패널조사에서 일반 사업장 부문에서 사용된 무응답 조정군(셀, cell)은 표본설계의 증화변수 중에서 지역구분을 무시하고 사업장의 업종과 사업장 규모를 고려하였다. 사업장의 업종과 사업장 규모가 동일하면 주요 조사변수가 유사한 값을 나타낼 것이고, 표본 사업장의 응답률도 유사할 수 있을 것이다. 응답여부(응답:1, 무응답:0)를 종속변수로 하고, 업종, 사업장 규모, 지역 등을 설명변수로 하는 로지스틱회귀모형(logistic regression model)을 적합한 결과에서도 업종과 사업장 규모 변수가 유의하게 나타났다.

공공부문 사업장은 전체를 하나의 층으로 간주하여 무응답 조정군으로 사용하였다. 표본조사에서 단위 무응답을 보정하기 위한 무응답 조정값(F_c : non-response adjustment)은 각 무응답 조정 군에서 다음 식에 따라서 구하였다.

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

여기서 n_1 은 조사가 가능 사업장 중에서 조사에 응답한 사업장 수이고, n_2 는 조사가 가능 사업장 중에서 조사에 응하지 않은 사업장 수이다.

무응답 조정을 거친 1차 가중치는 설계가중치와 무응답 조정값의 곱으로 구한다. 1차 가중치는 극단값 조정(weight trimming)의 단계를 거쳐 조정되었다. 1차 가중치는 해당 업종 구분과 사업체 규모에서 가중치의 중앙값을 3배 또는 1/3배를 넘지 않도록 조정되었다. 가중치의 극단값 조정은 사업체패널조

사의 추정결과에 약간의 편향(치우침)이 발생하게 할 수 있다. 반면에 가중치의 극단값을 조정함으로써 추정결과의 표준오차는 줄어든다. 결과적으로 가중치의 극단값 조정을 통해서 추정결과의 평균제곱오차(MSE: Mean Square Error)는 줄어들어 추정의 정확도가 높아질 것이다(Potter 1990; Kish 1992).

사후층화 조정은 통계분석에서 자주 사용될 것으로 판단되는 사업장 특성 변수에 대해서 모집단과 표본 자료의 분포를 일치시키는 과정이라고 할 수 있다. 사업체패널에서는 업종 구분(12)과 사업장 규모 구분(4)을 사후층화 변수로 이용하였다. 일반 사업장의 경우에 사후층화 조정을 위해서 사용된 자료는 사업체 패널의 조사시점과 거의 비슷한 사업체기초통계조사(2005년 12월 말 기준) 자료를 이용하였다.

2. 패널 가중치

2005 사업체패널조사는 1차년도 조사로 횡단면 조사의 성격만을 갖게 된다. 2차년도 조사부터는 횡단면 조사와 패널조사의 성격을 함께 지니기 때문에 횡단면 분석을 위한 가중치와 패널 가중치가 함께 산정되어 제공되어야 한다.

패널 가중치는 1차년도 횡단면 가중치와 2차년도 응답확률을 함께 이용하여 산정된다. 1차년도 패널 사업장 중에서 2차년도 패널에 응답할 확률은 표본 사업장 특성에 따라 차이가 날 것이다. 1차년도 조사에 응답한 표본 사업장이 2차년도 조사에 응답할 확률은 1차년도 조사결과의 사업장 특성변수를 설명변수로 하는 로지스틱회귀모형을 통해서 추정될 수 있다. 즉, 2차년도 패널에 응답했는지 여부를 반응변수로 하고, 산업분류, 지역, 근로자 수, 노조 유무, 설립연도, 본·지사 등의 변수를 설명변수로 하는 로지스틱회귀모형(Roberts et al. 1987)을 적합하여 응답률의 차이를 보정할 수 있을 것이다.

패널 가중치는 다음과 같이 구할 수 있다. 여기서 2차년도 조사에 응답할 확률은 앞서 설명한 로지스틱회귀모형을 통해서 추정한다.

$$\text{패널 가중치} = 1\text{차년도 가중치} \times \frac{1}{2\text{차년도 조사에 응답할 확률}}$$

V. 데이터 분석 및 이용

1. 데이터 분석과 설계요인

일반적으로 표본조사를 통해서 얻어진 데이터를 분석할 때는 표본설계에서 사용된 설계요인(예를 들면 층화, 집락, 가중치 등)을 반영하여 분석해야 한다. 사업체패널조사를 비롯한 복합표본조사 데이터를 분석할 때 표본설계와 가중치를 무시하고 분석하면 모수 추정에 심각한 편향(bias)이 발생할 수 있고, 추정량의 분산(표준오차)이 과소평가되어 문제가 될 수 있다. 모수 추정뿐만 아니라 신뢰구간 추정과 가설검정을 위한 유의확률(p-값) 산정에도 표본설계와 가중치를 반영해야 한다.

복합표본설계가 추정량의 분산에 미치는 영향은 설계요인(design factor: DEFT)이나 설계효과(design effect: DEFF)를 통해서 살펴볼 수 있다. 설계효과는 복합표본설계를 반영하여 구한 추정량의 분산과 단순임의추출을 가정하여 구한 추정량의 분산 비(ratio)이고, 설계요인은 설계효과의 양의 제곱근이다. 설계요인은 복합표본설계를 적용함으로써 같은 크기의 단순임의추출법에 비해서 얼마나 표준오차가 증가하게 되었는가를 나타낸다.

〈표 7〉은 사업체패널조사의 주요 질문항목에 대한 추정치와 표준오차 현황으로서 복합표본조사 분석용 통계패키지인 SUDAAN의 선형화 방법(Taylor method)에 의해서 계산한 결과이다. 사업체 대상의 조사에서 추정치의 표준오차를 계산해 보면 같은 표본크기의 단순임의추출법에 비해서 표준오차 값이 크게 나타난다. 결과적으로 설계요인을 계산해 보면 대부분 1보다 큰 값을 보이게 된다. 〈표 7〉을 보면 설계요인은 조사변수에 따라서 약간의 차이는 있지만 대략 1.3~1.5의 값을 갖고 있다. 단순임의추출을 가정하여 분석할 경우에 추정량의 분산(표준오차)이 과소평가되어 문제가 될 수 있음을 보여준다. 이기재(2001)와 Lee & Lepkowski(2001)은 우리나라의 소규모 사업체 대상의 조사에 대해서 설계효과와 가중치 효과를 분석하였다.

〈표 7〉 사업체패널조사의 주요 조사항목에 대한 추정값, 표준오차, 설계요인

조사 항목	추정값 (estimate)	표준오차 (s.e)	표본크기 (unweighted)	DEFT ⁸⁾
조직유형				
개인 사업장	7.0%	0.91	71	1.56
회사법인	79.6%	1.15	1411	1.24
학교법인 또는 의료법인	8.3%	0.74	167	1.17
회사 이외의 법인(학교, 의료 제외)	5.1%	0.75	256	1.49
여성근로자 비율	28.4%	0.72	1905	1.33
비정규직 근로자의 고용추세				
비정규직을 사용한 적이 없음	28.0%	1.51	351	1.47
비정규직 비중을 줄여 왔음	20.5%	1.31	469	1.42
비정규직 비중을 그대로 유지	34.4%	1.59	723	1.47
비정규직 비중을 늘려 왔음	8.5%	0.91	242	1.42
기타	8.7%	0.96	120	1.49
노동조합 조직 여부				
있다	19.9%	1.11	784	1.21
없다	79.4%	1.14	1112	1.23
휴면노조 상태	0.7%	0.34	9	1.75
전반적 노사관계				
매우 나쁘다	0.0%	0.01	3	0.30
나쁜 편이다	1.2%	0.35	29	1.43
보통이다	26.1%	1.49	461	1.48
좋은 편이다	58.2%	1.66	1126	1.47
매우 좋다	14.5%	1.17	286	1.45
교육훈련 실시 지원 여부				
있다	86.2%	1.21	1762	1.54
없다	13.8%	1.21	143	1.54
교육훈련 형태				
사내 직업훈련	38.1%	1.60	785	1.44
외부 직업훈련	39.7%	1.59	826	1.41
비공식적인 현장훈련	16.7%	1.28	239	1.49
입사 이전의 학교 교육 지원	0.4%	0.23	5	1.52
교육훈련이 필요 없음	5.1%	0.80	50	1.59

8) SUDAAN에서 제공하는 설계효과 계산 방법 중 두 번째 방법(DEF₂)에 따라 구한 결과이다.

2. 2005 사업체패널조사 자료 이용 안내

2005 사업체패널조사의 자료는 한국노동연구원 웹사이트(<http://www.kli.re.kr/wps>)에서 간단한 가입절차를 거쳐 다운받을 수 있다. 이때, (주)한국신용평가(한신평)의 재무정보를 포함한 데이터는 한국노동연구원의 출판실에 연락하여 신청서를 작성한 후 한신평 재무정보에 대한 비용을 지불하면 CD 형태로 받을 수 있다.

참고문헌

- 이기재. 2001. "복합표본조사 데이터 분석을 위한 회귀모형 접근법 비교 : 소규모사업체 조사 데이터 분석을 중심으로." 《조사연구》 2(1) : 73-86.
- Chaplin, J., Mangla, J., Purdon, S. and Airey, C. 2005. "The Workplace Employment Relations Survey (WERS) 2004." *Technical Report*.
- Kish, L. 1992. "Weighting for Unequal P_i ." *Journal of Official Statistics* 8: 183-200.
- Kish, L. 1987. *Statistical Design for Research*. New York : John Wiley & Sons.
- Lee, K. J. and Lepkowski, J. M. 2002. "Design and Weighting Effects in Small Firm Survey in Korea." *The Korean Communications in Statistics* 9(3): 775-786.
- Patak, Z., Hidiroglou, M. A. and Lavallée, P. 2000. "The Methodology of the Workplace and Employee Survey." *Proceedings of the Survey Research Methods Section*: 83-91. Alexandria, VA : American Statistical Association.
- Potter, F. 1990. "A Study of Procedures to Identify and Trim Extreme Sample Weights." *Proceedings of the Survey Research Methods Section*: 225-230. Alexandria, VA : American Statistical Association.
- Roberts, G., Rao, J. N. K., and Kumat, S. 1987. "Logistic Regression Analysis of Sample Data." *Biometrika* 74(1): 1-12.

[접수 2008/7/8, 1차 수정 2008/10/10, 2차 수정 2008/11/3,
 게재확정 2008/11/4]