

산업재해의 효율적 분석을 위한 표본설계*

황진수¹⁾

요약

산업재해는 일반재해, 중대재해, 직업병으로 나뉘어진다. 본 논문에서는 현재 산업안전공단에서 실시하고있는 일반재해에 관한 통계 산출방법을 검토하여 보고 효율적인 표본설계에 의한 새로운 통계 산출방법을 제시하고자 한다.

1. 서론

산업재해의 문제는 어느 나라에서나 매우 중요한 문제라고 생각하나 종종 잊고 지내는 경우가 많다. 그러나 보고되는 산업재해자 수만을 기초로 하더라도 평균적으로 우리 나라에서 산업재해로 인하여 작업장에서 매일 10여명이 목숨을 잃고 있고 매일 300여명 꼴로 부상을 당하거나 직업병에 걸리고 있다. 여기에서 위험한 작업환경에서 작업후 자신도 모르게 의복이나, 도구, 차량으로 인하여 가정까지 연결되어 가족들이 위험에 노출되는 것까지 고려한다면 문제는 더욱 심각해진다. 물론 이 수치는 5인 이상 산업재해보험에 가입한 사업장 중에서 보고되는 재해에만 의한 것이므로 실제 수치는 이보다 훨씬 클 것이라는 것이 일반적인 예상이다.

노동부에서 발행한 95년의 산업재해분석자료에 의하면 보고된 사망자 수는 2,662명이고 부상은 74,252건으로 나타나 있고 산재 보상금을 포함한 경제적 손실 추정액은 약 5조 6,000억을 상회하고 있음을 알 수 있다. 미국의 경우에는 1997년 NIOSH(National Institute for Occupational Injury and Safety)의 보고에 의하면 6,000건이 넘는 사망과 대략 200만건 이상의 부상건수가 있었으며 이로 인한 경제적 손실을 돈으로 추산해보면 약 1조 2,100억 불 정도가 된다고 보고하고 있다.

우리 나라의 경우 산재예방의 중요한 자료원천인 재해조사표의 작성이 92년부터 행정규제완화의 차원에서 권장사항으로 바뀐에 따라서 결과적으로 산재의 예방은 뒷전으로 물러났고 산재보험수령을 위하여 근로복지공단에 제출하는 요양신청서에 의한 사후 보상만이 이루어지고 있는 실정이다. 물론 요양신청서에도 재해에 관련된 항목을 포함하고 있다. 그러나 재해의 원인을 규명하기에 필요한 주요 항목들이 없기 때문에 요양신청서에 의한 재해통계분석은 명백한 한계를 가지고 있다.

산업재해 통계분석의 커다란 목적은 각 산업별로 주 위험요인을 도출하고 이에 따른 안전교육의 실시 또는 안전장치 등을 보완함으로써 산업재해를 줄이거나 예방하는데 있다고 볼 수 있다. 따라서 정확한 재해통계가 없다면 산업재해에 대한 올바른 제반정책의 결정이나 수립을 기대하기가 어려울 것이다. 예방 가능한 산재의 지속적인 발생이나 산재발생으

* 이 연구는 96년도 한국과학재단 연구비지원(과제번호: 96-0701-01-01-3)에 의한 결과임.

1) (452-751)인천시 남구 용현동 253, 인하대학교 통계학과, 부교수

로 인한 산재 보험율의 상승 등은 기업의 생산성을 저하시킬 뿐 아니라 간접비용의 상승으로 여러 측면에서의 기업의 경쟁력을 둔화시켜 결국은 국가의 대외 경쟁력을 약화시키는 요인으로 발전하게 된다.

현재 산업안전공단에서 수행하고 있는 재해자 표본조사는 산재보험 가입 사업장에서 발생한 산업재해(일반재해, 중대재해, 직업병) 중에서 근로복지공단에 의하여 4일 이상의 요양이 필요하다고 판정된 일반재해자의 요양신청서 리스트를 모집단으로 하여 이루어지고 있다. 현행의 표본조사는 요양신청서에서 파악하기 어려운 재해관련 세부 항목을 기존의 재해조사표를 이용하여 추출된 표본에 대하여 조사를 하는 것이다.

본 논문에서는 기존의 요양신청자 대상의 표본설계를 전 사업장(1인 이상)을 대상으로 한 표본설계 방법으로 전환하여 기존의 산업재해 표본설계가 가지고 있는 포함오차의 문제점을 해결하고 지역, 업종별 재해에 관한 통계를 산출할 수 있도록 하였다. 종전의 설계는 근로복지공단에서 제공되는 요양신청서를 모집단으로 하여 표본조사를 하였기 때문에 시간과 비용이 적게 드는 장점이 있었지만 제외되는 사업장과 산재은폐의 문제가 있어 정확한 산재의 실태를 파악하는데 어려움이 많았다. 이 조사가 제대로 시행이 된다면 모든 사업장에서 발생하는 재해에 관한 여러 가지 의미 있는 통계자료를 산출할 수가 있고 이를 바탕으로 산재보험 대상의 확대라든지 하는 정책적인 결정에 기초자료로 쓰일 수가 있을 것이다. 참고로 현재 산재보험 가입 대상은 종업원 5인 이상의 사업체를 대상으로 하고있다. 전 사업장대상의 표본설계는 일차적으로 재해 경험사업장을 찾아내고 차후에 재해 경험사업장에 대하여 다시 추출을 하는 이중표집법(double sampling)의 기법을 이용하여 진행을 한다.

2. 기존표본설계의 요약

산업재해 표본조사의 목적은 재해예방에 필요한 과학적이고 기술적인 산업재해통계를 생산 보급하는 것을 목적으로 하고 있다. 현재 일반적으로 작성되는 산업재해관련 통계는 재해보상을 위한 요양신청서에 의거하여 이루어지고 있는데 여기에서 실제 재해예방에 필요한 정보를 도출하기에는 부족하다고 할 수 있다.

현행 산업재해통계의 표본조사는 해당 년도의 상반기(1월1일 6월30일)중에 발생하여 요양이 결정된 재해자 리스트를 모집단으로 하여 이루어진다. 이 모집단의 리스트를 주요 5가지 항목인 지방노동관서(지역), 업종, 근로자수(규모), 직종, 요양일수 단계의 순서로 정렬하여 전체의 5%를 즉 1/20방식의 계통추출법을 이용하여 표본을 추출한다. 조사방법은 우편을 이용하여 재해조사표를 직접 작성하게 하는 자계식을 이용한다. 1차 무응답자에 한하여 2차 우편설문조사를 다시 수행하고 나머지 무응답자에 한하여 조사원이 직접 방문하여 면접조사를 수행하는 방식으로 최대한의 응답률을 확보하려고 한다. 이렇게 3차에 걸친 조사 후 최종 설문지 회수율은 대략 87%에 이르는 것으로 추정된다.

통계분석은 각 항목별로 도수분포표, 교차표, 그림 등을 이용하여 이루어지며 이러한 빈도분석을 통하여 모집단의 항목별 비율이나 총계 등을 추정하고 각 항목별로 신뢰구간을 제시한다. 현 조사에서는 응답자별 미응답 항목이 약 35%를 나타내어 응답내용을 이용

하여 효과적인 통계분석을 하기에 적지 않은 문제점을 가지고 있어 이에 대한 대책을 마련하고자 연구가 진행되고 있다.

3. 총화를 이용한 새로운 표본설계

3.1. 새로운 설계의 기본 원칙

전 사업체 대상 표본설계의 모집단은 산업안전보건법에 해당하는 모든 사업장(1인 이상)으로 현재 대한민국의 행정권이 미치는 전 지역에 소재 하는 모든 사업체를 조사 대상으로 한다. 단, 다음의 사업체는 제외한다.

- 가. 개인이 경영하는 농림 어업사업체(법인 및 비법인 단체가 경영하는 사업체는 조사대상)
- 나. 국방 및 가사서비스업
- 다. 국제기구 및 기타 외국기관
- 라. 고정설비가 없거나 영업장소가 일정치 않은 간이 판매상
- 마. 조사기준일 현재 개설 준비중이거나 3개월 이상 휴업상태인 사업체

전체 사업체에서 규모가 4인 이하인 사업체는 전체 사업체에서 85.79%(95년 기준)를 차지하고 있지만 산업재해보험 가입대상에서는 제외되어 있다. 새로운 표본설계의 모집단은 1996년 기준으로 통계청에서 조사한 사업체 기초통계조사에 참여한 모든 사업체가 될 것이다. 1995년의 사업체 기초통계조사자료를 보면 사업체 수는 2,771,068개이며 이는 1994년의 사업체수 2,518,454개에 비하여 252,614개가 증가한 것이다. 역동적인 경제구조와 급변하는 국내의 상황으로 인한 휴 폐업 및 새로운 벤처기업의 탄생 등으로 모집단의 변화가 적지 않으므로 이를 효과적으로 반영할 수 있는 새로운 표본설계가 필요하다. 또한 지방자치제의 실시에 따른 지역별 통계의 필요성이 어느 때 보다 중요하게 대두되고 있다.

본 논문에서는 이러한 다양한 요구를 반영할 수 있는 새로운 표본설계로서 다단계 층화 표본추출법을 제안하였다. 이러한 층화표본추출에 의한 설계는 각 층별 또는 세부 층(cell)별 통계 산출을 가능하게 할 뿐 아니라 이를 종합하여 전국적인 통계의 산출 및 추정을 가능하게 해준다. 기존의 재해자 대상 표본설계와 달리 사업체 대상 표본설계는 사업체가 추출 단위이기 때문에 사업체가 추출된 후에야 재해 발생 여부를 알 수가 있게 된다. 지금까지 보고된 자료를 분석하여 보면 사업체 당 산재발생 경험 비율은 그리 높지 않으므로 표본의 효율성을 높이기 위하여 이중표집법을 사용하게 된다. 즉 1차의 예비조사에서 산재발생 경험이 있는 사업체를 찾아낸 후 이를 부차모집단으로 간주하여 여기에서 다시 기존의 재해 조사표를 우편으로 발송하여 기존의 재해조사와 동일한 방법으로 표본조사를 수행하는 것이다.

새 표본설계는 소규모 사업체(1-4인)의 재해 빈도에 대한 일차적인 추정값을 구하고 기존 산재보험 가입대상 사업체(5인 이상)의 재해 빈도와 여러 각도에서 비교하는 것이 커다

란 목적 중의 하나라고 할 수 있다. 소규모 사업체에서의 재해 발생 형태나 업종의 분포가 5인 이상 사업체와는 다르리라고 예상된다. 소규모 사업체는 규모가 큰 사업체와는 달리 개 폐업 등의 변동이 심하여 표본의 관리나 추적도 어려우리라 예상된다. 따라서 설계 단계에서부터 소규모(1-4인), 5인 이상 사업체로 층화를 하여 독립적인 추정량을 얻을 수 있도록 설계를 한다. 산업재해 중에서 사망재해와 같은 중대재해나 직업병 등은 별도로 전수 조사가 이루어지므로 현재의 모집단에서는 제외되며 당해 년도의 일반재해자 전체가 모집단을 구성하게 된다.

3.2. 층화

층화의 기준은 재해분석에서 중요한 관심변수인 5 가지 항목, 지역(지방 노동관서), 업종, 근로자수(규모), 직종, 요양일수 등을 중심으로 층을 차례로 나누되 세분하게 되면 층에 배정되는 표본수가 적어지므로 정확한 추정을 할 수 없다.

새로운 표본설계에서는 세부적으로 모든 층별로 정확한 통계량을 구하는 설계를 제시하려는 것이 아니라 시 도별과 업종등 주요변수에 대하여 적절한 추론을 할 수 있는 설계를 제시하려는 것이다. 따라서 세부적인 조사는 별도의 계획에 의하여 필요한 표본수를 확보하여 시행하는 것이 바람직하다.

1 차 층화는 먼저 소규모사업체(1-4인)와 5인 이상 사업체 두가지로 하며 95년도의 사업체 기초통계조사에 의하면 사업체수의 비는 2,771,068 대 393,893으로 대강 소규모 사업체수가 7배정도 많은 것으로 나타나 있다.

2 차 층화는 각 층내에서 지역과 업종 등으로 세분해 나가는 것이다. 다음의 표는 개략적으로 층화를 하는 구조를 나타낸 것이다.

표 3.1: 전사업체 대상 표본 설계시 층화 구조표

1-4인				5인 이상			
서울특별시		...	제주도	서울특별시		...	제주도
광 업	...	기 타		광업 5-10인	
					...		
					...		

3.3. 표본크기의 결정 및 표본배정

표본설계에서 가장 먼저 결정되어야 할 것은 표본의 크기를 정하는 것이다. 표본의 크기를 결정하려면 먼저 주어진 조사에서 주 관심사가 되는 중요한 변수들을 선택하고 그 변수들을 표본으로부터 주어진 정도(precision)나 목표 변동계수(CV) 내에서 추정할 때 필요

한 최소한의 표본의 크기들을 구하고 그 중에서 가장 큰 것으로 결정하면 된다. 새로운 표본설계에서 주요변수가 모집단에서의 재해율이라 한다면 앞에서 언급한 것처럼 어느 정도의 적당한 오차의 한계 내에서 추정하기에는 표본의 크기가 너무 커지게 된다. 예를 들어 95년도의 경우를 보면 5인 이상의 산재보험 가입대상자 중에서 산재를 당한 모집단의 수가 $N = 78,034$ 명이며, 재해율은 0.99%(0.0099)로 발표되었다. 표본조사를 시행할 때 모집단 재해율을 목표변동계수(CV) 이내로 정확하게 추정하기 위한 표본수의 계산식은 다음과 같다.

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

여기서 CV는 목표정도이며 C는 모집단에서 재해율의 변동계수로서 가능하다면 보통 과거의 조사에서의 값으로 대체하여 이용한다. 즉, $\hat{C} = \sqrt{\frac{1-P}{P}} \times 100$ 로부터 $\sqrt{\frac{(1-0.0099)}{0.0099}} \times 100 = 10 \times 100 = 1,000(\%)$ 가 되어 전혀 신뢰할 수 없는 값이 된다. 통상적으로 추정량이 최소한의 신뢰성이 있으려면 CV의 값이 5~10% 정도는 되어야 한다. 따라서 CV=5%로 하여 위의 식에서 n을 구하면 26,445로 모집단 수 $N = 78,034$ 의 1/3정도나 되는 커다란 수가 되어 표본조사의 의미가 전혀 없게 된다. 따라서 만약 재해율의 추정값을 구할 수는 있는데 그 결과가 안정적이지 아니라면 추정을 하지 않는 것이 좋다는 결론에 도달한다.

새로운 표본설계는 사업체가 추출단위인 새로운 추출방법이므로 표본 크기의 결정은 예산이나 조사 인력, 시간등 외적인 변수에 의하여 결정을 하게된다. 과거의 조사에 소요된 시간 및 비용 등을 토대로 하여 연간 2회의 표본조사를 한다고 가정하여 최종 통계분석을 하게 되는 것은 분기별 2,000 건 정도의 자료를 분석한다고 정하였다. 이 수치는 기존 조사에 비하여 약간 많은 수치이나 소규모 사업장에 대한 조사도 포함되므로 그리 많은 수는 아니다. 이 수치를 기반으로 하여 일차 표본의 크기등이 결정되며 이는 조사 비용에 따라 영향을 받게된다.

위의 수치는 최종 분석을 원하는 수치이므로 이중표집에서 첫단계에 필요한 표본수는 일차 응답지 회수율과 재해사업장적중율을 역산하여 구하게 된다. 응답지 회수율은 기존조사에서의 회수율(3차에 걸친) 87%를 근거로 하고 재해사업장적중률은 95년의 산업재해분석 자료를 토대로 하면 186,021개의 사업장에서 한 번이라도 산재가 발생하여 신고한 사업장수는 34,801개로서 18.7%가 된다. 은폐되는 산재를 고려한다면 이보다 더 많은 비율을 예상할 수 있으나 소규모 사업장은 산재의 발생 비율이 높은 제조업과 건설업체의 수가 5인 이상 사업장에 비하여 32.57% 대 10.6%로 적기 때문에 산재발생이 적을 것이라고 예상된다. 따라서 전체적으로 재해 발생사업장적중율을 15%로 잡고서 역산을 해보면 일차 조사에서 필요한 표본수는 대략 30,000건이 된다.

새로운 설계의 목적 중의 하나는 처음 시행하게되는 소규모 사업체에서의 각종 재해의 양상을 기존의 5인 이상 사업체에서의 재해와 비교(예를 들어 업종별 재해율의 차이등)에 있으므로 일단 동일한 크기의 표본을 배정하는 동등배정을 사용하여 각 층에 15,000개의 표본을 배정한다.

다음으로 지역과 업종간에는 상관관계가 있기 때문에 지역 내에서 업종을 따로 층화 한다면 전국적으로 볼 때 업종의 균형이 맞지 않게 될 가능성이 높아 전국적으로 지역과 업종을 동시에 고려하는 설계를 하여야 한다. 따라서 각 규모별(4인 이하 와 5인 이상)로 층

화변수로서 지역과 업종을 동시에 고려하여 배정된 15,000개의 표본을 전국 각 시도 15개와 산업대분류 업종(광업, 전기.가스.수도업, 제조업, 운수.창고.통신, 건설업, 기타) 6가지를 고려하여 $15 \times 6 = 90$ 개의 층으로 나누어 각 층에 배정을 한다. 이 때 각 셀에 배정하는 표본수는 지역별 재해사업장적중률과 업종별 재해사업장적중률을 동시에 고려하여 최적배정을 하게 된다.

지역을 중심으로 먼저 배정을 하면 다음과 같다. 모집단이 L개의 지역(층)으로 층화가 되어있고 지역별 사업체수가 각각 N_1, N_2, \dots, N_L 이라 하면 전국 사업체의 총수는 $N = N_1 + \dots + N_L$ 이 된다. 만약 각 지역에서 재해를 경험한 사업체의 모비율을 각각 $p_i, i = 1, \dots, L$ 라 한다면 전국에서 재해를 경험한 사업체의 모비율(p_{st})은 다음과 같이 추정된다.

$$\widehat{p}_{st} = \frac{1}{N}(N_1\widehat{p}_1 + \dots + N_L\widehat{p}_L) = \frac{1}{N} \sum_{i=1}^L N_i\widehat{p}_i$$

추정된 \widehat{p}_{st} 의 분산추정량은 다음과 같이 계산된다.

$$\widehat{V}(\widehat{p}_{st}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{\widehat{p}_i \widehat{q}_i}{\widehat{n}_i - 1} \right)$$

이 때 n_i 는 i번째 지역에서 추출된 표본 사업체 수를 나타내며 $\widehat{q}_i = 1 - \widehat{p}_i$ 는 i번째 지역에서의 재해를 경험하지 않은 사업체의 표본비율을 나타낸다.

각 지역(층)마다 한 개의 표본을 추출하는데 드는 비용(c_i)까지 고려하여 최적배정된 표본의 크기 n_i 는 소요비용(c_i)의 제곱근에 반비례한다. 통상적으로 우편조사 등에서는 각 층마다 자료를 얻기 위한 비용이 동일하므로, 즉 $c_i = c$ 라 할 수 있으므로 각 층별로 네이만 배정법에 의하여 배정되는 표본의 수는 $N_i \sqrt{p_i q_i}$ 에 비례하게 된다. 이 때 p_i 는 모집단에서의 값이므로 시험조사(pilot survey)의 결과 또는 과거의 자료나 정보를 이용한 추정값(\widehat{p}_i)을 사용하게 된다.

이 추정값 \widehat{p}_i , 즉 i번째 지역의 재해사업장적중률은 97년 상반기에 조사한 산업재해분석에서 재해자 자료를 사업장과 지역별로 구별하여 지역 전체의 사업장 수로 나눈 값으로 사용을 한다.

마찬가지로 업종별 재해사업장적중률을 고려하여 위와 같은 방법으로 업종별로 표본배정을 하여 지역별 배정 수와 조정을 하여 최종 층별로 배정되는 표본수를 결정하게 된다.

다음은 네이만배정에 필요한 사업장 수에 관한 표들로서 표 3.2는 규모와 업종별 사업장수이며 표 3.3은 규모와 지역별 사업장 수이다.

표 3.2: 규모, 업종별 사업체수의 비('95 사업체 기초통계조사)

	광업	제조업	전기 가스 수도	건설업	운수 창고 통신	기타	계
1-4인 (비율)	1,556 (0.065)	212,349 (8.93)	412 (0.017)	40,288 (1.69)	194,514 (8.18)	1,928,056 (81.118)	2,377,175 (100)
5인이상 (비율)	1,175 (0.30)	101,934 (25.88)	842 (0.21)	26,563 (6.74)	16,911 (4.29)	246,468 (62.58)	393,893 (100)

표 3.3: 규모, 지역별 사업체수의 비('95 사업체 기초통계조사)

	사업체 수		계
	1 - 4 인 (%)	5 인이상 (%)	
서울특별시	598,104 (25.2)	109,921 (27.9)	708,025
부산광역시	219,942 (9.3)	36,426 (9.3)	256,368
대구광역시	143,973 (6.1)	22,476 (5.7)	166,449
인천광역시	106,457 (4.4)	20,213 (5.1)	126,670
광주광역시	67,361 (2.8)	10,082 (2.6)	77,443
대전광역시	71,520 (3.0)	10,970 (2.8)	82,490
경기도	335,017 (14.1)	65,264 (16.6)	400,281
강원도	88,430 (3.7)	12,234 (3.1)	100,664
충청북도	73,858 (3.1)	11,018 (2.8)	84,876
충청남도	94,223 (4.0)	13,174 (3.3)	107,397
전라북도	99,405 (4.2)	13,565 (3.5)	112,970
전라남도	109,293 (4.6)	14,715 (3.7)	124,008
경상북도	137,639 (5.8)	19,054 (4.8)	156,693
경상남도	203,430 (8.6)	29,586 (7.5)	233,016
제주도	28,523 (1.2)	5,195 (1.3)	33,718
전지역	2,377,175	393,893	2,771,068

다음 표 3.4는 네이만배정으로 실시한 표본배정의 결과를 보여주고 있다.

표 3.4: 네이만배정법에 의한 표본배정의 결과

	광업	제조업	전기 가스 수도	건설업	운수 창고 통신	기타	전산업
서울특별시	4	1,751	1	413	194	2,722	5,085
부산광역시	1	474	1	111	52	736	1,375
대구광역시	1	232	0	55	26	361	675
인천광역시	1	275	0	65	30	428	799
광주광역시	0	73	0	17	8	114	212
대전광역시	0	87	0	21	10	135	253
경기도	2	1,103	1	260	122	1,715	3,203
강원도	0	104	0	25	11	162	302
충청북도	0	87	0	21	10	135	253
충청남도	0	113	0	27	12	176	328
전라북도	0	115	0	27	13	179	334
전라남도	0	101	0	24	11	157	293
경상북도	0	178	0	42	20	278	518
경상남도	1	454	0	107	50	706	1,318
제주도	0	18	0	4	2	28	52
전국	10	5,165	3	1,219	571	8,032	15,000

위의 표에서 보듯이 네이만배정의 결과 광업과 전기·가스·수도업의 경우는 각각 10개와 3개의 사업체만 배정되는 결과를 낳았다. 현재의 조사에서 중점을 두어야 할 것은 업종별 재해율이므로 실제로 조사를 할 경우에는 정책적으로 기타 사업장을 분리하여 예를 들어 5,000개의 사업장을 배정하고 나머지에 10,000 사업장을 다시 배정하여 전체적으로 광업과 전기·가스·수도업에 조사할 만한 표본 사업장이 배정되도록 조정하여야 할 것이다.

사업장 표본을 다시 세부적인 층으로 배정 할 때 각 층마다 최소표본과 최대표본의 수나 비율의 하한선과 상한선을 정하는 것이 사후 표본의 관리 등의 측면에서 바람직하다. 예를 들어 어느 지역의 어떤 업종에 배정된 표본수가 3일 때 업종의 변화가 심한 경우에는 조사를 못하게 되는 경우도 생길 것이다. 따라서 층별 최소 배정 표본수를 5로 하며 최대는 전체의 반을 넘지 않도록 한다는 원칙을 세워 이를 지켜나가는 방법으로 하는 것이 바람직할 것이다. 또한 배정된 표본수를 5의 배수로만 한다든지 하는 조정은 표본의 사후 관리상 여러 가지로 편리하다. 이처럼 표본 크기의 결정이나 표본의 배정은 이론적으로 보면 여러 방법이 제시되고 있으나 현실적인 비용이나 조사원의 수 또는 지역적인 문제 그리고 사후 표본의 관리 등의 문제까지 고려하며 조정을 하여 결정하는 것이 바람직하다.

4. 표본추출 및 추정

4.1. 1단계 표본추출 및 추정

가) 표본추출틀의 준비

표본추출틀은 1996년 12월 31일 기준으로 조사한 사업체 기초통계조사 및 도소매업 서비스업 총조사 결과의 전사업체를 시·도별, 산업 대분류별, 사업체 규모(종사자수)별로 분류한 사업체 명단을 작성하여 표본추출틀을 구성한다.

실제로 이 조사를 수행하기 위해서는 통계청과 산업안전공단의 적극적인 협조가 필요하다. 위의 자료는 통계법 상에 엄격하게 규제되어 있으므로 이 중에서 조사목적에 부합하는 항목에 대한 정보를 얻으려면 통계청의 자료제공 지침에 따라 산업안전공단의 책임자로부터의 공식적인 자료협조 요청이 이루어져야 한다. 이 때 현재 시행하려는 표본조사의 목적을 자세하게 설명하여야 할 것이다.

나) 표본추출

모집단의 리스트를 일차적으로 사업체 규모에 따라 1-4인과 5인 이상의 두 층으로 나눈다. 소규모 사업체(1-4인)의 리스트를 지역에 따라 정렬하고 지역 내에서 다시 산업 대분류별로 정렬을 한다. 이 후에 앞에서 배정한 표본수 만큼을 계통추출법에 따라 표본을 추출한다.

5인 이상의 사업체 리스트에서도 역시 지역에 따라 정렬하고 다음에 업종×규모에 따라 정렬한 후 미리 정해진 수만큼의 표본을 계통추출법에 의하여 추출한다. 이상과 같이 추출된 표본에 대하여 1단계의 예비설문을 보낸다.

다음 단계는 예비설문을 보낸 곳에 전화를 걸어 협조를 당부하며 응답율을 높이기 위해 노력한다. 이렇게 하여 결과적으로 회수된 설문지 결과를 일차적으로 분석하여 원하는 정보를 얻는다. 이 중에서도 재해가 한 번이라도 있었다고 응답한 사업장은 별도로 리스트를 작성한 다음 이를 지역, 업종, 규모 등에 따라 정렬하여 2단계 표본조사를 위한 리스트를 구성한다.

다) 추정

전 사업장에서 재해사업체의 비율을 각 층 별로 추정을 하고 이를 종합하여 시·도별, 산업 대분류별, 규모별 재해사업체 비율을 추정한다.

예를 들어 경기도지역(i지역)에서 제조업(j업종)분야의 소규모 사업체(1-4인)의 총수를 N_{ij} 라하고 여기에 배정된 표본 수를 n_{ij} 라 하자. 이 n_{ij} 개의 사업체 중에서 회신을 한 사업체 수를 n'_{ij} 라하고 이 중에서 재해가 한번이라도 발생하였다고 응답한 사업체 수를 x_{ij} 라 하면 경기도의 소규모 제조업에 종사하는 사업체 중 한 번이라도 재해를 경험한 사업체의 비율(p_{ij})의 추정 값은 $\widehat{p}_{ij} = \frac{x_{ij}}{n'_{ij}}$ 가 된다.

이 추정값의 분산 추정량은 다음과 같다.

$$\widehat{Var}(\widehat{p}_{ij}) = (1 - f_{ij}) \left(\frac{\widehat{p}_{ij}\widehat{q}_{ij}}{n'_{ij} - 1} \right)$$

단 f_{ij} 는 해당 층의 표본추출율로서 $\frac{n_{ij}}{N_{ij}}$ 이다. 만일 경기도 지역의 모든 업종에 대하여 재해 경험 사업체 비율(p_i)을 추정하고 그 분산을 추정하려면 다음과 같이 가중평균을 이용한다.

$$\hat{p}_i = \sum_{j=1}^L \frac{N_{ij}}{N_i} \hat{p}_{ij}$$

$$\widehat{Var}(\hat{p}_i) = \sum_{j=1}^L \left(\frac{N_{ij}}{N_i}\right)^2 \widehat{Var}(\hat{p}_{ij})$$

여기에서 $N_i = \sum_j N_{ij}$ 로서 경기지역의 모든 사업체 수를 말한다.

4.2. 2단계 표본추출 및 추정

1단계에서 재해경험사업장으로서 부차모집단을 형성한 후 여기에서 다시 표본을 추출하는데 이 때는 각 사업체마다 재해경험 횟수를 가중치로 하여 사업체를 추출한다. 이는 상대적으로 재해가 많이 발생한 사업장을 표본에 포함 시킴으로서 재해원인 분석에 도움을 주고 또한 많은 재해자의 자료를 얻을 수 있으므로 효율적인 방법이라 할 수 있다. 처음 예상대로 결과가 집계된다면 2단계에서는 전체에 대하여 다 조사하여야만 최종분석을 원하는 2,000개 정도의 표본이 될 것이다. 이론적으로 2단계 추출은 미리 정한 표본의 개수를 주요변수인 지역, 업종, 규모 등의 순서대로 정렬한 후 재해발생 횟수를 고려하는 크기비례확률표본추출법을 사용하는 것이 바람직하다. 현재의 설계에서 2단 추출을 하려면 1단계에서의 표본수가 더 많아져야 의미가 있다.

2단계에서 표본조사의 목적은 기존의 요양신청서에서 알 수 없었던 여러 항목에 대한 정보를 얻으려는 것이다. 즉, 재해의 발생형태, 기인물, 작업내용, 보호장비, 작업공정, 상해부위, 상해종류 등에 관한 추가적인 통계분석을 함으로써 재해의 원인과 예방에 도움이 되는 의미 있는 통계량을 산출하고자 한다. 이 경우에 추정대상 모집단은 사업체가 아니라 전체 산업재해(4인 이하 사업체 포함)자가 된다. 즉 사업체별로 재해발생 건수에 따라 가중치를 주었으므로 결국 모든 재해자가 모집단이 되는 것이다.

추정은 최종적으로 회수되어진 재해조사표를 근거로 하여 이루어진다. 추정하는 방법은 결과적으로 앞서의 방법과 동일하므로 여기에서는 생략한다. 단 여기에서도 소규모 사업체(1-4인)와 5인 이상 사업체와는 별도로 추정을 하고 전사업체 관련 통계량이 필요하면 앞서와 같이 합하여 통계량을 산출하면 된다.

5. 결론 및 추후과제

새로운 표본설계는 일차적으로 소규모 사업장에 대한 재해 관련한 정보를 처음으로 제공하게 될 것이며 사업장대상의 재해를 조사와 기존의 재해자대상 표본조사와의 비교를 통하여 보다 정확한 산재 관련 통계를 산출할 수 있으리라 여겨진다. 전사업장대상 표본설계가 성공적으로 수행된다면 산재보상보험 적용근로자의 범위를 모든 근로자로 확대하기 위한 중요한 기초자료가 되리라고 생각한다.

다단계 층화설계에서 앞으로 고려해야할 사항으로는 현재 선진국에서 점차로 고려하고 있는 자영업자에 관한 부분의 추가가 될 것이다. 이는 1997년의 NIOSH 심포지엄에서도 별도의 섹션으로 이 부분을 다룬 것을 보면 알 수 있을 것이다.

또 하나의 문제는 업종 중에서 기타항목에 관한 부분이다. 기타(농·임·서비스업 등)업종이 차지하는 비중은 1995년말의 자료에 의하면 소규모(1-4인)업체인 경우 전체 사업체 수에서 81.12%이며 5인 이상 사업체에서는 62.58%로서 사업체를 대상으로 추출할 때 상당히 많은 부분을 차지하고 있으나 재해율로 본다면 97년 상반기 자료에 의하면 11.62%에 불과하다. 따라서 이 기타부분을 전체에서 분리하여 별도로 관리하는 것이 표본설계의 효율을 높일 수 있을 것으로 생각된다.

참고문헌

- [1] 노동부(1995), <산업재해분석>.
- [2] 산업안전공단(1996), <주요국의 산업재해현황 및 통계제도>.
- [3] 산업안전공단(1997), <산업재해 표본조사 분석>.
- [4] 전홍석, 김승환(1996), <SAS를 활용한 윈도우 프로그래밍>. 자유아카데미.
- [5] 통계청(1996), <사업체기초통계조사 보고서(전국편)>.
- [6] Cochran(1977). *Sampling Techniques, 3rd edition*. New York, Wiley.
- [7] Kish, L.(1965). *Survey Sampling*. New York, Wiley.
- [8] Little, R. and Rubin, D.(1987). *Statistical Analysis with Missing Data*. New York, Wiley
- [9] Scheaffer, R., Mendenhall, W., and Ott, L.(1996), *Elementary Survey Sampling, 5th edition*. Duxbury Press.

[1998년 9월 접수, 1999년 5월 최종수정]

A New Sampling Design for the Industrial Hazards and Casualty Statistics

Jinsoo Hwang¹⁾

ABSTRACT

Currently statistics on the national workplace injuries and illnesses are compiled by the Korea Industrial Safety Corporation(KISCO). There are two types of injuries, one is minor injury and the other is fatal(major) injury. In this study we have analyzed the current statistical system which generate the injury statistics and try to suggest an efficient sampling design in order to generate an enhanced injury statistics.

1) Associate Professor, Department of Statistics, Inha University, 402-751, Incheon, Korea