

중소기업실태조사를 위한 표본설계

김달호¹ · 황진섭² · 곽상규³

¹²³ 경북대학교 통계학과

접수 2010년 8월 1일, 수정 2010년 8월 1일, 게재확정 2010년 10월 10일

요약

중소기업의 조사부담 완화를 위하여 2009년부터 기존의 “중소기업실태조사”, “중소기업인력실태조사” 및 “중소기업사업전환실태조사”를 통합하여 하나의 조사로 수행하게 되었다. 통합 중소기업실태조사를 위한 새로운 표본설계에서 중소제조업과 지식기반서비스업 부문이 너무 상이하므로 별개의 모집단을 구성하였으며, 전체 표본수는 10,000개로 하였다. 변동계수를 사용한 배분공식으로 1차 층인 종사자규모·산업분류별로 표본을 배분하고 루트비례배분을 사용하여 2차 층인 지역별로 표본의 크기를 결정하였다. 새로운 표본설계에서는 가중치를 계산하였고 이를 이용한 추정량과 추정오차 공식을 유도하여 기존의 단순집계를 벗어나 종사자 규모별, 업종별 그리고 지역별 추정과 추정의 정도에 대한 평가를 가능하게 하였다.

주요용어: 가중값, 계통추출, 루트비례배분, 목표오차, 층화.

1. 서론

중소기업통계는 중소기업의 활동현황, 자금, 인력 및 경영실태 등에 대한 내용을 조사하여 이들의 현황과 변동추이를 정확하게 파악함으로써 정부 및 유관기관의 중소기업 지원정책 수립과 구조개선에 필요한 기초자료로 제공된다. 중소기업의 부담완화를 위하여 2009년부터 기존의 “중소기업실태조사”, “중소기업인력실태조사” 및 “중소기업사업전환실태조사”를 통합하여 하나의 조사로 수행하게 되었다. 통합 중소기업실태조사에서는 기존 조사의 통합에 따라 가능한 조사 일관성은 그대로 유지하고 주요 시계열 자료에 대한 체계적 관리를 도모하고자 한다. 뿐만 아니라 조사부담의 경감을 위해 조사항목을 간소화하고 응답자의 편의성을 고려하여 설문을 배열하고 조사의 신뢰성 제고를 위해 주관적 판단에 좌우되는 설문을 최소화하고자 한다.

매년 실시되어온 기존의 3개 조사가 통합됨에 따라 기존 조사의 시계열 유지를 도모하면서 전년도 조사와 연계된 통합 통계표작성이 요구되므로 통합 조사에 대한 효율적 조사 방안을 강구할 필요성이 대두되어 2009년도 중소기업실태조사를 위한 표본설계에 관한 연구를 수행하게 되었다. 특히 중소제조업의 경우 기존 실태조사에서 표본 추출틀로 사용되던 통계청의 “광업·제조업 통계조사”가 2008년에는 5-9인 규모에 대해 표본조사로 대체되고 2009년도부터는 5-9인 규모에 대한 조사가 폐지됨에 따라 새로운 추출틀에 대한 연구가 필요하게 되었다. 표본추출방법과 표본규모를 결정하고 업종별, 규모별, 지역별 표본을 배분하기 위해 새로운 모집단의 특성을 분석하였다. 이를 위해 SAS 프로그램을 사용하였다 (SAS, 2009). 이 때 중소제조업과 지식기반서비스업 부문이 너무 상이하므로 별개의 모집단을 구성하였다. 모집단 분석을 바탕으로 표본설계를 새롭게 개편함으로써 모집단을 잘 대표하는 표본을 선정할 수 있게 되었다.

¹ 교신저자: (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 교수.
E-mail: dalkim@knu.ac.kr

² (450-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 박사.

³ (450-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 박사과정.

2. 현행 조사에 대한 분석

지금까지 실시되었던 기존의 “중소기업실태조사”, “중소기업인력실태조사” 및 “중소기업사업전환실태조사”를 살펴봄으로 통합 조사에서 고려해야 할 여러 가지 사항을 검토하였다. 특히 기존의 조사 중 “중소기업실태조사”와 “중소기업인력실태조사”를 표본설계 측면에서 살펴보았다.

2.1. 중소기업실태조사

중소제조업의 2007년 실태전반을 파악하여 중소기업 지원 및 구조개선에 필요한 기초자료 제공하기 위하여 중소기업실태조사가 실시되었다. 조사내용으로 경영일반, 공장보유, 구매, 생산설비 (기계장치) 투자, 판매, 수·위탁 거래, 고용, 재무구조 등의 항목을 중점적으로 조사하였다.

표본설계는 기업체를 추출단위로 하고 통계청의 2006년 기준 광업·제조업 통계조사 결과를 이용하여 한국 표준산업 중분류별 (20개), 종사자 규모별 (6개)로 층화한 중소기업체를 추출틀로 사용하였다. 층화기준은 20개 업종별 (산업중분류 2자리수)로 1차 층화하고, 이를 6개 종사자 규모별로 2차 층화하였다.

표본수는 네이만 최적배분법에 의해 층내 부가가치를 주요 변수로 사용하여 신뢰수준 95%, 목표정도 (요구분산) 5.9%로 산업중분류별 (20개), 종사자규모별 (6개) 표본수를 결정하였다. 표본추출 방법으로는 층화단순임의추출법을 사용하였다. 또한, 표본조사를 근거로 산출된 모수 추정치의 신뢰성을 평가하기 위해 표본오차를 고려하였다. 4,100개 업체를 표본 조사하여 수집된 결과 (표본치)를 표본설계시 정해진 확대비율 (모집단수/표본수)을 적용하여 모수를 추정하였고 모수를 추정하였다.

2.2. 중소기업인력실태조사

2.2.1. 제조업 부문

제조업 부문 표본설계에서는 산업대분류상 제조업 (중분류 15~37) 가운데 5인 이상 300인 미만 사업체 (단, 제조업 가운데 담배 제조업 (16)은 제외)를 대상으로 하였다. 또한, 2005년 기준 통계청 “광업·제조업통계조사”상의 종업원 수 5인 이상 300인 미만의 제조업을 표본 추출틀로 사용하였다. 표본의 크기는 8,000개로 설정하였고, 95% 신뢰구간 하에서 전국 산업소분류 기준 허용오차는 5%, 지역별 (16개 시·도) 산업중분류 기준 허용오차 8%로 설정하였다.

표본설계에서는 인력실태조사에서 각 종사자 규모별 특성을 살펴보기 위해 종사자 규모에 따른 층화를 실시하였다. 16개의 시·도로 층화하고 각 시·도의 22개의 중분류업종을 1차층 (부차모집단)으로 간주하고 각 중분류업종에서 4개의 종사자 규모층 (2차층)이 존재하는 2원 층화표본추출을 수행하였다. 16개 시·도에서 22개 중분류업종 각각을 모집단으로 설정하여 루트비례 할당하였다. 그러나 세세분류 부모집단의 크기가 작을 경우, 전수 할당하였다.

2.2.2. 서비스업 부문

서비스업의 표본설계는 지식기반 서비스업으로 분류될 수 있는 19개 서비스업종을 영위하는 중소기업체를 대상으로 하였다. 또한, 2005년 기준 통계청 “도소매업 및 서비스업 총조사” 보고서 상의 서비스업 19개 업종을 표본 추출틀로 사용하였다. 표본의 크기는 2,000개로 설정하였고, 95% 신뢰구간 하에서 전국 산업소분류 기준 허용오차는 5%, 지역별 (16개 시·도) 산업중분류 기준 허용오차 8%로 설정하였다.

표본설계는 인력실태조사에서 각 종사자 규모별 특성을 살펴보기 위해 종사자 규모에 따른 층화를 실시하였다. 34개의 세세분류업종을 1차층 (부차모집단)으로 간주하고 각 세세분류업종에서 4개의 종사

자 규모층(2차층)이 존재하는 2원 층화표본추출을 수행하였다. 이 때 34개 세세분류 업종 각각을 모집단으로 설정하여 루트비례 할당하였다. 여기서 세세분류 부모집단의 크기가 작을 경우, 전수 할당하였다. 또한, 자연과학 연구개발업은 종사자 규모별 3개 층으로 층화하고, 시장조사 및 여론조사업, 경영상담업 등 일부 업종에 대해서는 종사자 규모별 층화에서 4규모는 100~199인의 범위로 정의하였다.

3. 새로운 표본설계

3.1. 새로운 표본설계의 특징

통합 중소기업실태조사를 위한 새로운 표본설계에서 전체 표본수는 조사의 일관성과 조사의 시계열 유지를 위해 기존의 “중소기업인력실태조사”에서 사용되었던 중소기업 부문 8,000개와 지식기반서비스업 부문 2,000개를 합한 총 10,000개의 표본을 그대로 유지하기로 하였다. 조사에서 표본 사업체는 종전과 같이 사업체 전체와 종사자규모별, 사업체별, 지역별, 종사자규모·사업체별, 종사자규모·지역별, 사업체별·지역별 그리고 주요 항목에 대한 통계작성에 중점을 두었다.

업종별·종사자규모별로 1차 층화를 하였고, 16개 시도별로 2차 층화를 하였다. 표본 사업체의 배분은 종사자 규모의 목표오차를 고려한 표본할당 공식을 사용하여 1차 층에 배분하였다. 그리고 각 1차 층별로 루트비례배분을 사용하여 2차 층에 배분하였다(김영원 등, 1998). 각 층별 사업체 목록을 작성해서 조사를 수행하기 전에 사업체의 변동을 반영하여 추가로 예비표본조사구를 추출하였다. 새로운 표본설계에서는 가중치 계산과 이를 이용한 추정량은 물론 추정오차 공식을 유도하였다.

3.2. 모집단 분석

새로운 표본설계에서 중소기업 부문에 조사모집단은 산업대분류상 제조업 가운데 종사자수가 5인 이상 300인 미만인 전국의 기업체이며, 조직형태가 회사이외의 법인, 국가·지방단체, 비법인단체인 경우와 모집단의 규모가 작아 통계자료로서의 의미가 없는 담배제조업 경우는 조사 모집단에서 제외하였다. 중소기업 경우 조사모집단으로 기존의 조사에서 사용되었던 통계청의 “광업·제조업조사” 자료를 검토하였으나, 이는 2008년에는 5-9인 규모에 대해 한시적으로 표본조사로 대체되었다가 2009년도부터는 5-9인 규모에 대한 조사가 폐지됨에 따라 새로운 조사모집단에 대한 연구가 필요하게 되었다. 대안으로 통계청의 “전국사업체조사” 자료를 새로운 조사모집단으로 검토하였으나, 법인 코드 및 사업자 등록번호에 대한 낮은 응답률로 인해 시도별 서로 다른 지역에 산재해 있는 본사와 공장 혹은 한 공장과 다른 공장 자료의 파악이 불가능하게 되어 기업체를 추출단위로 사용하지 못하는 문제가 발생하게 되었다. 따라서 통계청의 “광업·제조업조사” 자료에서 중소기업체의 절반 정도에 해당하는 5-9인 규모 업체에 대한 모집단 자료 보완이 절실하게 되었다.

종사자수가 5-9인 규모의 업체는 대부분 사업체와 기업체 구분이 크게 차이가 없을 것으로 예상되므로 이를 동일한 것으로 가정할 수 있다. 만약 실제 자료를 사용하여 이 가정이 성립함을 확인할 수 있다면 종사자 5-9인 업체에 대해서는 “전국사업체조사” 자료를 사용하고 10인 이상 규모의 업체는 “광업·제조업조사” 자료를 사용하여 새로운 조사모집단으로 사용할 수 있다. 종사자수가 5-9인 규모의 경우 2006년도와 2007년도에 걸쳐 “광업·제조업조사”와 “전국사업체조사”의 두 모집단을 비교 검토한 결과 거의 동일한 분포를 가지는 것으로 확인할 수 있었다. 따라서 종사자 5-9인 업체에 대해서는 “전국사업체조사” 자료를 사용하고 10인 이상 규모의 업체는 “광업·제조업조사” 자료를 사용하여 새로운 조사모집단을 구성하였다.

3.3. 층화

새로운 표본설계는 통합 중소기업실태조사에서 각 업종별, 규모별, 지역별 통계작성 뿐만 아니라 세부 항목인 업종 및 종사자 규모별 추정이 가능하도록 층화하였다. 중소기업은 총 6개의 종사자규모별과 총 23개 업종의 산업중분류별로 1차 층화하였다. 따라서 제조업의 1차 층화는 총 138개 (6×23)의 층으로 구성된다. 그리고 총 16개의 시·도별로 2차 층화하였다. 지식기반서비스업은 총 4개의 종사자규모별과 총 46개 업종의 산업세분류별로 1차 층화하였다. 따라서 서비스업의 1차 층화는 총 184개 (4×46)의 층으로 구성된다. 그리고 총 16개의 시·도별로 2차 층화하였다.

3.4. 표본크기 결정 및 배분

새로운 표본설계에서 전체 표본수는 중소기업 부문 8,000개와 지식기반서비스업 부문 2,000개를 합한 총 10,000개의 표본을 유지하기로 하였으며, 업종별, 규모별, 지역별 추정치들에 대한 표본오차 뿐만 아니라 세부 발표항목인 업종 및 종사자 규모별 추정에 대한 표본오차 관리를 위해 업종·종사자 규모별 목표오차 설정하여 1차 층의 표본크기를 결정하였다. 추정량의 정도는 각 업종별 모집단 사업체 수 규모에 따라 표본오차의 절대량으로 목표오차를 설정하기 보다는 상대표본오차 개념인 변동계수 (Coefficient of Variation; CV)를 사용하는 것이 바람직하다고 판단하여 변동계수를 사용하여 각 층의 목표오차를 설정하였다. 그리고 표본설계에서 각 층의 표본크기는 모집단의 중요변수인 종사자수를 고려하여 결정하였다.

구체적인 목표오차는 제조업의 경우 각 산업중분류의 각 기업체 수와 지식기반서비스업의 경우 각 산업세분류의 각 사업체 수에 따라 다르게 하는 것이 바람직하다. 중소기업에서는 기업체 수가 1000개 이하의 산업중분류에 대해서는 목표오차를 5.0%를, 기업체수가 1001개에서 5000개 이상 산업중분류에 대해서는 목표오차를 2.5%를, 기업체수가 5001개 이상 산업중분류에 대해서는 목표오차를 2.0%를 사용하여 표본크기를 결정하였다. 한편, 지식기반서비스업에서는 사업체 수가 100개 이하의 산업세분류에 대해서는 목표오차를 9.5%를, 사업체수가 101개 이상에서 1000개 이하의 산업세분류에 대해서는 목표오차를 7.5%를, 사업체수가 1001개 이상의 산업세분류에 대해서는 목표오차를 5.5%를 사용하여 각 층의 표본크기를 결정하였다.

이렇게 정해진 목표오차를 충족시키기 위해서는 각 층을 독립적인 모집단으로 보고 각 층별로 요구되는 목표정도에 따라 표본의 크기를 결정해야 한다. 제조업과 서비스업 모두에 대해 업종 (제조업: 산업중분류, 서비스업: 산업세분류)·종사자 규모별 표본수는 다음의 공식에 따라 결정되었다 (이기재 등, 1997).

$$n = \left(\frac{C}{C_{\bar{y}}} \right)^2 / \left[1 + \frac{1}{N} \left(\frac{C}{C_{\bar{y}}} \right)^2 \right] \quad (3.1)$$

여기서 C 는 모집단 변동계수로 $C = S/\bar{Y}$ 이고, $C_{\bar{y}}$ 는 목표오차 그리고 N 은 업종별·종사자 규모별 모집단 업체수이다. 표3.1과 표3.2는 제조업 산업중분류별·종사자규모별 모집단 및 표본 현황과 서비스업 산업세분류별·종사자규모별 모집단 및 표본 현황을 각각 나타낸다.

한편, 층의 표본을 결정할 때 업종·규모별 층의 모집단 사업체수의 크기가 작아 표본 추출이 안 된 경우는 전수할당을 하였다. 만약 각 지역별로 세부 발표항목인 업종 및 종사자 규모별 추정에 대한 표본오차를 관리하고자 한다면 지역표본의 수가 크게 늘어 조사가 사실상 불가능하다. 따라서 각 지역별·종사자 규모별 통계작성 및 표본오차를 관리가 가능하도록 표본수를 결정하였다. 지역별 표본수는 각 업종·규모별 층에 할당된 표본에서 각 16개 시·도에 루트비레 할당법인 다음의 공식으로 표본을 할당하였

표 3.1 중제조업의 산업중분류별·종사자규모별 모집단 및 표본 현황

산업중분류	종사자 규모											전체		
	5~9인		10~19인		20~49인		50~99인		100~199인		200~299인			
	모집단	표본	모집단	표본	모집단	표본	모집단	표본	모집단	표본	모집단	표본		
10_식료품	3968	108	1803	106	1118	167	330	82	130	56	37	20	7386	539
11_음료	159	14	120	15	58	19	23	11	13	8	1	1	374	68
13_섬유제품 (의복제외)	3968	108	1877	109	1017	160	245	71	87	44	21	14	7215	506
14_의복/모피제품	4969	106	2132	103	811	146	170	65	73	39	18	11	8173	470
15_가죽/가방,신발	1084	67	465	62	218	71	41	25	14	11	5	4	1827	240
16_목재/나무제품 (가구제외)	1377	69	570	61	210	77	19	14	11	8	0	0	2187	229
17_펄프/종이/종이제품	1580	71	839	65	448	93	103	41	47	27	14	8	3031	305
18_인쇄/기록매체복제업	2285	69	965	70	359	87	47	22	26	17	2	2	3684	267
19_코르크/연탄/석유정제품	83	14	63	12	34	13	11	6	1	1	1	1	193	47
20_화학물질/화학제품 (의약품제외)	1910	68	1106	66	643	104	184	50	95	42	29	14	3967	344
21_의료용물질/의약품	148	14	105	13	108	22	79	13	55	13	10	2	505	77
22_고무제품/플라스틱제품	4216	108	2686	108	1647	176	384	79	154	59	46	18	9133	548
23_비금속광물제품	2178	69	1194	68	694	102	119	42	68	32	13	9	4266	322
24.1차금속제조업	1391	64	1007	65	724	108	213	47	106	36	31	14	3472	334
25_금속가공제품 (기계/가구제외)	9660	107	4962	107	2543	177	458	85	169	61	43	15	17835	552
26_전자부품/컴퓨터/영상/음향/통신장비	2556	106	1716	107	1296	165	435	86	269	71	81	21	6353	556
27_의료/정밀/광학기기/시계	2062	70	959	64	503	97	109	34	67	31	20	12	3720	308
28_전기장비	3765	108	1969	108	1124	159	301	82	121	54	27	15	7307	526
29_기타기계/장비	7852	110	4644	107	2622	181	633	89	221	73	48	19	16020	579
30_자동차/트레일러	1447	66	1118	65	1102	113	401	53	190	46	80	15	4338	358
31_기타운송장비	408	57	328	59	369	79	245	53	159	43	29	11	1538	302
32_가구	2058	67	902	68	421	93	50	29	21	14	4	3	3456	274
33_기타제품	1884	70	578	60	262	77	47	23	24	17	2	2	2797	249
총합계	61008	1710	32108	1668	18331	2486	4647	1102	2121	803	562	231	118777	8000

다 (Cochran, 1997).

$$n_{hij} = n_{hi} \times \frac{\sqrt{N_{hij}}}{\sum_{j=1}^{16} \sqrt{N_{hij}}} \tag{3.2}$$

여기서, n_{hi} 는 각 업종·규모별 표본수, n_{hij} 는 각 업종·규모·지역별 표본수, N_{hij} 는 각 업종·규모·지역별 모집단 업체수이다.

3.5. 표본 추출

중소제조업의 경우 표본추출틀은 2007년도 기준으로 통계청에서 조사한 “광업·제조업 통계조사” 결과 중 종사자수 10인 이상 300인 미만의 전체 기업체와 “전국사업체조사” 자료 중 종사자수 5-9인 전체 기업체를 결합하여 기업체 명단을 작성하여 사용하였다. 단, 조직형태가 회사이외의 법인, 국가·지방단체, 비법인단체인 경우와 모집단의 규모가 작아 통계자료로서의 의미가 없는 담배제조업 경우는 제외하였다. 지식기반서비스업의 경우 표본추출틀은 2007년도 기준으로 통계청에서 조사한 “전국사업체조사” 결과 중 종사자수 5인 이상 300인 미만의 전사업체 명단을 작성하여 사용하였다. 여기서도 조직형태가 회사이외의 법인, 국가·지방단체, 비법인단체인 경우는 제외하였다. 표본추출틀에서 업종·규모별, 지역별 층화를 한 후 각 층에 표본을 할당하고, 할당된 표본의 크기대로 각 층에서 표본을 추출하기 위해 적절한 분류변수를 선정하여 분류변수에 따라 업체를 정렬 후 계통추출법으로 표본을 추출하였다. 이 때 분류변수로 조직형태, 사업체구분, 행정구역번호 (시군구, 읍면동), 종사자수, 창업년월 등을 고려하였다.

표 3.2 지식기반서비스업의 산업세세분류별·종사자규모별 모집단 및 표본 현황

산업세세분류	종사자 규모								전체	
	5~19인		20~49인		50~99인		100~299인		모집단	표본
	모집단	표본	모집단	표본	모집단	표본	모집단	표본		
39001_토양 및 지하수 정화업	4	4	1	1	1	1	1	1	7	7
39009_기타 환경 정화 및 복원업	10	7	1	1	0	0	0	0	11	8
58211_온라인·모바일 게임 소프트웨어 개발 및 공급업	157	25	69	13	29	5	20	8	275	51
58219_기타 게임 소프트웨어 개발 및 공급업	72	24	27	9	5	3	5	3	109	39
58221_시스템 소프트웨어 개발 및 공급업	1793	54	619	25	193	10	97	22	2702	111
58222_응용 소프트웨어 개발 및 공급업	1002	55	294	24	86	13	35	16	1417	108
59111_일반 영화 및 비디오물 제작업	100	22	12	6	6	4	1	1	119	33
59112_애니메이션 영화 및 비디오물 제작업	93	21	30	8	9	5	3	3	135	37
59113_광고 영화 및 비디오물 제작업	193	24	32	7	2	2	0	0	227	33
59114_방송 프로그램 제작업	189	25	38	9	15	6	0	0	242	40
59120_영화, 비디오물 및 방송프로그램 제작 관련 서비스업	27	11	5	5	2	2	0	0	34	18
61210_유선통신업	214	20	257	11	130	3	83	14	684	48
61220_무선통신업	119	21	35	9	12	6	9	5	175	41
61230_위성통신업	2	2	1	1	0	0	0	0	3	3
61291_통신 제관매업	98	24	30	6	9	5	7	4	144	39
61299_그외 기타 전기 통신업	75	22	27	8	9	5	9	4	120	39
62010_컴퓨터 프로그래밍 서비스업	549	31	88	16	16	4	7	5	660	56
62021_컴퓨터시스템 통합 자원 및 구축 서비스업	251	25	98	9	37	8	30	12	416	54
62022_컴퓨터시설 관리업	88	23	18	7	6	4	1	1	113	35
62090_기타 정보기술 및 컴퓨터운영 관련 서비스업	56	15	12	5	3	2	2	2	73	24
63111_자료 처리업	49	15	16	5	7	4	3	3	75	27
63112_호스팅 및 관련 서비스업	37	15	5	3	4	3	3	2	49	23
63120_포털 및 기타 인터넷 정보매개 서비스업	122	28	38	10	15	3	14	7	189	48
63991_데이터베이스 및 온라인정보 제공업	249	31	77	12	25	5	8	4	359	52
70111_물리, 화학 및 생물학 연구개발업	107	25	24	9	7	3	6	4	144	41
70112_농학 연구개발업	30	13	5	2	2	2	3	2	40	19
70113_의학 및 약학 연구개발업	114	21	18	7	6	4	6	5	144	37
70119_기타 자연과학 연구개발업	10	6	3	3	1	1	0	0	14	10
70121_전기·전자공학 연구개발업	217	25	71	12	23	4	22	7	333	48
70129_기타 공학 연구개발업	172	26	58	9	19	7	18	10	267	52
71400_시장조사 및 여론조사업	93	23	41	10	8	4	11	8	153	45
71531_경영컨설팅업	821	32	117	14	25	6	18	7	981	59
72121_건물 및 토목엔지니어링 서비스업	963	54	337	24	95	15	54	23	1449	116
72122_환경컨설팅 및 관련 엔지니어링 서비스업	136	30	29	12	4	2	1	1	170	45
72129_기타 엔지니어링 서비스업	936	53	260	22	78	12	34	19	1308	106
72911_물질성분 검사 및 분석업	132	20	43	9	8	3	3	3	186	35
72919_기타 기술 시험, 검사 및 분석업	742	21	188	10	40	5	14	8	984	44
72921_측량업	261	29	21	11	7	5	5	4	294	49
72922_제도업	60	15	8	5	7	2	0	0	75	22
72923_지질조사 및 탐사업	58	8	9	4	2	2	0	0	69	14
72924_지도제작업	36	13	17	9	4	3	2	2	59	27
73201_인테리어 디자인업	211	24	12	7	1	1	0	0	224	32
73202_제품 디자인업	208	25	16	7	4	3	0	0	228	35
73203_시각 디자인업	271	29	30	9	1	1	0	0	302	39
73209_기타 전문 디자인업	71	15	6	4	2	2	1	1	80	22
75992_전시 및 행사 대행업	527	24	97	10	27	5	11	6	662	45
75994_포장 및 증진업	194	24	59	11	32	5	11	7	296	47
85504_온라인 교육 학원	72	21	22	8	11	5	3	3	108	37
총합계	11991	1120	3321	438	1035	205	561	237	16908	2000

지식기반서비스업의 경우에도 계통추출을 위해 각 층에 속한 사업체를 동일하게 분류하여 정렬하였다. 한편, 실제 조사를 수행했을 때 표본 기업체 (사업체) 중에서 표본으로 사용할 수 없는 기업체 (사업체)가 발생하는 경우에는 교체 표본을 사용하는데 이를 위해 예비표본을 선정한다. 예비표본은 표본 기업체 (사업체)로 선정된 기업체 (사업체)를 제외하고 나머지 기업체 (사업체)들을 대상으로 필요한 수의 예비표본을 표본 기업체 (사업체)의 선정과 동일한 계통추출방법으로 추출하여 사용한다. 예비표본은 실제표본 (제조업: 8,000개, 서비스업: 2,000개)의 2배수만큼 선정하였다. 우선 모집단의 사업체수가 작아서 전수조사를 한 경우에는 예비표본은 없다. 두 번째로 모집단의 사업체수에서 표본의 사업체수를 제외한수가 표본의 사업체수의 두 배 보다 작은 경우 모집단의 사업체수에서 표본의 사업체수를 제외한 모든 사업체를 예비표본으로 선정하였다. 마지막으로 모집단의 사업체수에서 표본의 사업체수를 제외한 수가 표본의 사업체수의 두배보다 큰 경우인 경우 모집단의 사업체수에서 표본의 사업체수를 제외한 사업체 중에서 표본의 사업체수의 2배에 해당하는 수만큼 예비표본을 추출하였다. 이때 사용한 추출방법은 실제 표본을 추출한 방법인 업종·종사자규모·지역별 계통추출을 통하여 예비표본을 추출하였다. 예비표본은 제조업인 경우 13,980개, 서비스업인 경우 2,961개로 선정되었다.

4. 추정

중소기업실태조사 자료로부터 조사항목에 대한 업종별, 규모별, 지역별 추정값과 세부항목인 업종 및 종사자 규모별 추정값을 얻고자 한다. 뿐만 아니라 추정값과 더불어 표준오차 및 상대표준오차 공식을 제시하였다. 먼저 추정식에서 사용하게 될 기본적인 기호들을 정의하면 x_{hijk} 는 h 업종분류내 i 규모내의 j 지역에서의 k 번째 표본 업체의 해당 조사항목의 조사값, N_{hij} 는 h 업종분류내 i 규모내의 j 지역에서의 모집단 업체수, n_{hij} 는 h 업종분류내 i 규모내의 j 지역에서의 표본 업체수, $w_{hij} = N_{hij}/n_{hij}$ 는 가중치, $s_{hij}^2 = \frac{1}{n_{hij} - 1} \sum_{k=1}^{n_{hij}} (x_{hjik} - \bar{x}_{hij})^2$ 는 h 업종분류내 i 규모내의 j 지역에서의 표본분산, $\bar{x}_{hij} = (\sum_{k=1}^{n_{hij}} x_{hjik}) / n_{hij}$ 는 h 업종분류내 i 규모내의 j 지역에서의 표본평균, $h = 1, \dots, H$ 는 업종분류를 나타내는 첨자, $i = 1, \dots, I$ 는 종사자 규모를 나타내는 첨자, $j = 1, \dots, 16$ 는 지역 (시·도)를 나타내는 첨자, $k = 1, \dots, n_{hij}$ 는 (h, i, j) 층 내의 k 번째 업체를 나타내는 첨자를 나타낸다. 여기서 업종분류는 중소기업의 경우 산업중분류를 일컫고 지식기반서비스업의 경우 산업세분류를 일컫는다. 각 업종분류별 추정값과 추정값에 대한 분산은 다음과 같이 계산된다.

$$\widehat{T}_h = \sum_{i=1}^I \sum_{j=1}^{16} w_{hij} \sum_{k=1}^{n_{hij}} x_{hijk}; \quad Var[\widehat{T}_h] = \sum_{i=1}^I \sum_{j=1}^{16} N_{hij}^2 \left(\frac{N_{hij} - n_{hij}}{N_{hij}} \right) \cdot \frac{s_{hij}^2}{n_{hij}} \quad (4.1)$$

업종분류별 총계 추정값에 대한 표준오차와 상대표준오차는 다음과 같이 계산된다.

$$SE(\widehat{T}_h) = \sqrt{Var[\widehat{T}_h]}; \quad CV[\widehat{T}_h] = \frac{SE(\widehat{T}_h)}{\widehat{T}_h}. \quad (4.2)$$

각 종사자 규모별 총계 추정값과 추정값에 대한 분산은 다음과 같이 계산된다.

$$\widehat{T}_i = \sum_{h=1}^H \sum_{j=1}^{16} w_{hij} \sum_{k=1}^{n_{hij}} x_{hijk}; \quad Var[\widehat{T}_i] = \sum_{h=1}^H \sum_{j=1}^{16} N_{hij}^2 \left(\frac{N_{hij} - n_{hij}}{N_{hij}} \right) \frac{s_{hij}^2}{n_{hij}}. \quad (4.3)$$

종사자 규모별 총계 추정값에 대한 표준오차와 상대표준오차는 다음과 같다.

$$SE(\widehat{T}_i) = \sqrt{Var[\widehat{T}_i]}; \quad CV[\widehat{T}_i] = \frac{SE(\widehat{T}_i)}{\widehat{T}_i}. \quad (4.4)$$

지역별 총계 추정값과 추정값에 대한 분산은 다음과 같이 계산된다.

$$\widehat{T}_j = \sum_{h=1}^H \sum_{i=1}^I w_{hij} \sum_{k=1}^{n_{hij}} x_{hijk}; \quad Var[\widehat{T}_j] = \sum_{h=1}^H \sum_{i=1}^I N_{hij}^2 \left(\frac{N_{hij} - n_{hij}}{N_{hij}} \right) \frac{s_{hij}^2}{n_{hij}}. \quad (4.5)$$

지역별 총계 추정값에 대한 표준오차와 상대표준오차는 다음과 같다.

$$SE(\widehat{T}_j) = \sqrt{Var[\widehat{T}_j]}; \quad CV[\widehat{T}_j] = \frac{SE(\widehat{T}_j)}{\widehat{T}_j}. \quad (4.6)$$

만약 업종별, 규모별, 지역별 평균 추정값과 세부항목인 업종 및 종사자 규모별 평균 추정값을 구하고자 한다면 위에 주어진 총계 추정값 공식에서 가중치의 합을 나누어주면 되고, 평균 추정값의 분산은 가중치 합의 제곱을 나누어 주면 된다.

5. 표본관리 및 표본 대체

중소기업실태조사의 대상이 되는 종사자수 5인 이상 9인 미만의 기업체 (사업체)는 중소기업제조업의 경우 61,008개로 모집단의 51.4%를 차지하고, 지식기반서비스업의 경우 11,991개로 모집단의 70.9%를 차지하는 것으로 나타났다. 이렇게 소규모 기업체 (사업체)가 차지하는 비중이 매우 높는데, 소규모 기업체 (사업체)는 특성상 창업, 폐업, 휴업, 그리고 전업이 경기변동에 민감하므로 정확한 중소기업실태 조사를 위해서는 모집단과 표본의 변화에 따른 관리가 매우 중요하다.

소규모 중소기업의 경우 시간이 흐름에 따라 휴·폐업, 사업체 규모의 변동 등과 같이 기업체 (사업체) 특성의 변화가 심하다. 따라서 모집단의 변화를 반영하는 대표성 있는 표본을 사용하기 위해서는 모집단의 변동을 계속 파악해야 하고 표본을 적절히 교체해 주어야 한다. 중소기업실태조사에 있어서 모집단의 변화는 향후 휴·폐업, 규모의 변동, 급속한 산업 발달에 따른 산업의 질적·양적 변화로 인하여 발생할 수 있다. 모집단의 변동에 따른 해당 층의 표본의 증가나 감소는 해당 층에서 추출물의 역수만큼 변동이 생길 경우에 모집단 변동에 비례하여 증가 또는 감소시킨다.

표본으로 뽑힌 기업체 (사업체)에 변동이 생겨 표본의 대체가 불가피한 경우는 소멸되는 기업체 (사업체)가 속한 층에서 사업체를 선정하여 새로운 표본으로 선정하도록 한다.

중소기업실태조사의 모집단으로 사용되는 통계청의 “광업·제조업 통계조사”와 “전국사업체조사”는 매년 실시되고 있어, 새로운 기업체 (사업체)가 생기거나, 기업체 (사업체)가 없어지는 경우에 이를 파악하여 모집단을 매년 수정할 수 있다. 특히 공단조성과 같은 급격한 모집단의 변동이 있을 때에는 표본에 새로 생긴 사업체들을 반영시켜야 한다.

참고문헌

- 김영원, 류제복, 박진우, 홍기학 (1998). <표본조사의 이해와 활용>, 자유아카데미, 서울.
 이기재, 전중우 (1997). 노동통계조사를 위한 표본설계-매월노동통계조사, 노동력수요동향조사를 중심으로-. <응용통계연구>, 10, 215-226.
 Cochran, W. G. (1977). *Sampling techniques*, Wiley, New York.
 SAS (1999). *SAS/STAT user's guide, version 8*, SAS Institute Inc., Cary.

A sample design for the survey on actual state of SMEs

Dal Ho Kim¹, Jin Seub Hwang², Sang Gyu Kwak³

¹²³Department of Statistics, Kyungpook National University

Received 1 August 2010, revised 1 August 2010, accepted 10 October 2010

Abstract

In 2009 the former three surveys on small and medium enterprises were combined into one survey to reduce the response burden of enterprises. In this report, we study a sample design for 2009 survey on actual state of small and medium enterprises (SMEs). However, the differences between the manufacturing industries and knowledge-based service industries are so large that we need to consider separate populations in the survey. The total sample size is decided as 10,000 in new survey design for integration. We allocate the sample sizes for the first stratum based on CV and then allocate the sample sizes for the second stratum using root proportional formula. Also we calculate survey weights and propose the formula for the estimators as well as standard errors using weights for each characteristic.

Keywords: Coefficient of variation, root proportional allocation, stratification, systematic sampling, weights.

¹ Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

² Doctor of philosophy, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.

³ Doctor of philosophy student, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.