

부품·소재산업 동향 조사의 표본설계†

남궁 평¹⁾

요약

2006년 통계청이 시행한 광업·제조업 통계조사 결과(2005년 실적치)를 모집단으로 사용하면 최근 동향을 반영한 표본설계가 가능하다. 본 논문은 기존의 12개 업종보다 세분화된 94개 세부업종에 분류에 따라 매월 부품·소재산업의 생산, 출하, 재고의 변동사항을 조사하여 부품·소재산업의 경기변동실태를 파악하고 부품·소재산업의 육성정책 및 기업경영의 기초자료를 제공할 수 있는 새로운 표본설계를 제안한다. 표본설계는 응용절사법과 주성분을 이용한 다변량 네이만 배정법을 이용하여 층별로 표본크기를 결정하여 배정하고 표본추출은 확률비례계통추출을 사용한다.

주요용어: 표본설계; 표본 오차; 응용절사법; 확률비례계통추출법.

1. 서론

기존에 실시된 부품·소재산업 동향조사는 2003년 통계청에서 시행한 광업·제조업 통계조사 결과(2002년 실적치)를 모집단으로 하여 표본설계되었으므로 모집단의 변화를 반영하지 못하고 있다. 현재 기업체들이 부도, 폐업, 업종변경, 창업, 해외 이전 등 많은 변동이 발생하고 있어 소규모에서 대규모 또는 대규모에서 소규모로의 구조적 변동이 있으므로 이를 반영한 대표성 있는 표본을 추출하여 표본의 현실 반영도를 높이고 표본조사의 질적 향상을 위해 표본개편이 필요하다.

기존의 표본설계는 부품·소재 산업 12개 분류 업종별로 층화하여 추출했기 때문에 94개 세부업종의 부품·소재 산업의 정보를 추정할 때 편이가 발생하거나 추정효율이 떨어질 가능성이 높다. 따라서 94개 세부업종별로 적정한 표본오차를 유지할 수 있는 표본설계를 실시하고 세부업종별 부품·소재 산업 동향을 파악할 수 있는 정보를 추정 가능하도록 표본개편을 제안한다.

부품·소재산업 동향조사는 업종분류(12개 업종)와 세부업종분류(94개 업종) 수준에서 매월 부품·소재산업의 생산, 출하, 재고의 변동사항을 조사함으로써 부품·소재산업의 경기변동실태를 파악하여 부품·소재산업 육성정책 및 기업경영의 기초자료를 제공하는데 목적이 있다. 총 3개의 조사 항목인 생산부문, 출하부문, 재고부문에 대해 조사한다. 생산부문에서는 생산수량, 생산금액의 2개 조사 내용을, 출하부문에서는 내수출하수량, 내수출하금액, 수출출하수량, 수출출하금액의 4개 조사 내용을, 재고부문에서는 재고수량, 재고금액

† 이 논문은 2007년도 성균관대학교 성균학술연구비 지원에 의하여 연구되었음.

1) (110-745) 서울시 종로구 명륜동 3가 53, 성균관대학교 통계학과, 교수. E-mail: namkung@skku.edu

표 1.1: 한국표준산업분류에 따른 업종명

한국표준산업분류	업종명
17류	섬유제품제조업
24류	화학물 및 화학제품제조업
25류	고무 및 플라스틱제품 제조업
26류	비금속광물제품 제조업
27류	제 1 차 금속산업
28류	조립금속제품 제조업
29류	기타 기계 및 장비제조업
30류	컴퓨터 및 사무용기기 제조업
31류	기타 전자기계 및 전기변화장치 제조업
32류	전자부품, 영상, 음향 및 통신장비 제조업
33류	의료, 정밀, 광학기기 및 시계제조업
34류	자동차 및 트레일러 제조업
35류	기타 운송장비 제조업

의 2개 조사 내용을 조사한다. “부품·소재 전문기업 등의 육성을 위한 특별 조치법” 시행령 제 6조 ‘별표1’의 규정에 의해 한국표준산업분류 상 표 1.1에 해당하는 업종 중 부품·소재를 생산하는 사업체 중 표본사업체를 모집단으로 한다. 여기서 34류 자동차 및 트레일러 제조업, 35류 기타 운송장비 제조업은 통합하여 수송기계부품 업종으로 구분한다.

조사 주관기관은 한국기계산업진흥회 (2007)이며, 조사방법은 매월 조사를 실시하고 자계식을 원칙으로 하되 타계식(우편조사, 서면조사, 전화조사, 팩스조사, 방문조사, E-mail 조사, CASI(Computer Assisted Self-Interviewing) 등)을 혼용한다.

2. 모집단 분석

2.1. 모집단 설정

부품·소재산업 동향조사를 위해 모집단은 다음과 같이 정의한다.

1. 목표모집단: “2006년 광업·제조업 통계조사” 결과를 기초로 생산액이 0보다 큰 업종 [17, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35(KSIC2단위)] 중 “부품·소재 특별법 시행령” 제 6조 ‘별표 1’의 규정에 의한 부품·소재를 생산하는 사업체
2. 조사모집단: 목표모집단 중 94개 세부업종별로 생산액 누계 95% 사업체이며, 생산액 누계가 95%로 정하기 어려운 경우 생산액 누계 95% 이상 되는 최소 사업체 수로 조사모집단을 선정

모집단 정보는 2006년 광업·제조업 통계조사의 결과를 이용하고 세부업종별 구분하여 표본을 설계한다. 목표모집단과 조사모집단의 총 크기는 표 2.1과 같다.

2006년도에 실시된 광업·제조업 통계의 원시자료(2005년 실적치)를 이용하여 주요 항목별 특성을 분석한다. 즉, 주요 항목별 모평균과 표준편차, 변동계수(CV)등을 계산하여

표 2.1: 세부업종별 조사모집단 정의

(단위 : 백만원, 개 사업체)

업종	세부업종	절사점	총액대비비율(%)	조사모집단	목표모집단
섬유제품	제사 및 방직	1,175	95.02	289	499
	직물제조	1,063	95.01	599	961
	섬유염색및가공	1,626	95.03	169	275
	기타 섬유제품	964	95.02	331	576
화합물 및 화학제품	석유화학계 기초 유기화합물	35,250	96.61	4	7
	기타 기초 유기 유도체 및 화합물	56,011	95.53	36	102
	기타 기초 무기 화합물	3,502	95.02	88	208
	무기안료, 염료, 유연제 및 기타 착색제	1,548	95.08	146	235
	합성고무 및 플라스틱 물질	43,619	95.18	53	172
	기초의약품질 및 생물학적 제제	2,455	95.16	38	81
	농약	8,368	95.91	16	41
	도료 및 인쇄잉크	2,371	95.01	185	379
	계면활성화	1,645	95.17	77	121
	사진용화학제품 및 감광재료	1,285	95.00	39	85
	그외 기타 화학제품	1,357	95.00	523	912
고무 및 플라스틱 제품	화학섬유	10,574	95.09	31	90
	고무타이어 및 튜브	66,354	96.32	8	31
	제1차플라스틱제품	1,267	95.01	877	1,507
	기계장비 조립용 플라스틱 제품	1,107	95.00	1,082	2,020
비금속 광물제품	기타고무제품	932	95.00	163	277
	관유리	800	99.84	3	4
	산업용 유리제품	1,537	95.02	200	645
	산업용도자기	570	95.27	39	62
	구조용 정형내화제품	3,266	95.67	6	21
	석회	192	97.87	10	12
제1차 금속제품	기타 비금속광물제품	700	95.03	191	301
	합금철	4,107	96.11	8	17
	철강압연, 압출 및 연신제품	26,255	95.04	60	319
	철강관	2,390	95.00	319	632
	비철금속 제련, 정련 및 합금	8,448	95.07	66	199
	비철금속압연, 압출 및 연신제품	3,768	95.01	220	471
	철강주조업	1,314	95.01	133	227
조립 금속제품	철강주조업(절단가공)	25,187	95.34	14	72
	금속탱크, 저장조 및 유사용기	933	95.04	250	471
	원자로 및 증기발생기부품	385	95.04	36	51
	분말야금, 금속단조, 압형제품	846	95.00	1,030	1,690
	금속열처리업, 도금 및 기타처리업	498	95.00	1,109	1,488
	톱 및 호환성공구	620	95.02	236	394
	금속파스너 및 스프링	719	95.01	644	1,026
일반 기계부품	기타 금속제 부품	1,044	95.08	86	186
	내연기관 및 터빈	2,086	95.04	108	323
	펌프 및 압축기	960	95.01	445	776
	텀, 밸브 및 유사장치	817	95.03	453	716
	베어링, 기어 및 동력전달장치	886	95.00	318	601

업종	세부업종	절사점	총액대비비율(%)	조사모집단	목표모집단
	산업용오븐 및 노용부품	611	95.06	41	61
	운반하역기계부품	687	95.04	235	351
	냉동공조기계부품	927	95.01	869	1,508
일반 기계부품	가공공작기계부품	491	95.03	255	344
	농업용기계부품	423	95.08	167	251
	금속주조 및 기타 야금용 기계 부품	772	95.24	26	46
	건설광산기계부품	1,046	95.04	180	322
	음식료품기계 및 담배가공기계부품	165	95.98	15	22
	섬유기계 부품	285	95.07	152	192
	반도체제조장비부품	938	95.01	219	339
	무기 및 총포탄 부품	16,044	95.65	22	65
	가정용기기부품	742	95.01	310	528
	금형	592	95.00	1,819	2,557
	기타 일반기계부품	499	95.04	490	678
	컴퓨터 및 사무기기 부품	컴퓨터 기억장치	29,186	95.58	4
컴퓨터 입력장치		432	95.22	51	80
컴퓨터 단말기		801	95.60	31	41
컴퓨터용 메인보드 및 카드		856	95.18	18	36
컴퓨터 주변기기 및 전용부품		1,045	95.06	118	227
사무용기기부품		1,004	95.01	39	62
전기 기계부품	전동기, 발전기, 전기변환장치	1,196	95.00	692	1,380
	전기공급 및 전기제어장치	850	95.00	1,338	2,271
	절연선 및 케이블	1,330	95.03	284	584
	절연광섬유케이블	36,949	95.03	6	25
	일차전지 및 축전지	7,580	95.32	26	89
	진구 및 램프부품	748	95.20	26	63
	조명장치 부품	379	95.22	65	95
	기타 전기장비 부품	2,049	95.03	239	579
전자부품	개별소재 반도체	1,939	95.07	94	209
	집적회로 반도체	86,942	95.41	27	138
	통신기기 및 방송장비	1,126	95.01	577	1,162
	전자관	10,100	95.82	5	23
	인쇄회로판	3,270	95.05	169	451
	전자축전기	3,132	95.16	38	108
	전자저항기	1,538	95.25	31	71
	전자카드	826	95.05	29	51
	액정표시장치	49,714	95.06	45	224
	그외 기타 전자부품	1,095	95.01	216	443
정밀 기계부품	방사선장치 및 전기식 진단기기부품	450	95.14	17	30
	기타의료용기기 및 부품	348	95.21	65	103
	측정, 시험, 항해, 기타정밀기기 및 부품	734	95.02	477	800
	광섬유 및 광학요소	2,269	95.13	53	162
	시계부품	280	95.53	24	35
수송 기계부품	자동차용 엔진	33,000	97.68	5	8
	자동차부품	2,752	95.00	1,390	3,293
	선반구성부분품	1,416	95.03	130	225
	철도차량부품 및 관련장치	900	95.23	58	98

업종	세부업종	절사점	총액대비비율(%)	조사모집단	목표모집단
	항공기용 엔진 및 부품	3,949	95.34	10	48
	모터사이클 부품	389	95.23	44	67
	자전거부품	100	95.95	10	13
	합계			21,989	39,877

표본규모와 표본배정에 사용한다. 또한 기존의 표본설계 전 과정을 상세히 검토해서 현행 조사의 문제점, 개선내용 등을 종합적으로 살펴서 새로운 표본설계에 반영한다.

2.2. 모집단 분석

기존 표본은 12개 업종별 추정을 목적으로 표본을 설계하여 표본추출을 실시하였다. 현실적으로 이 표본을 바탕으로는 94개 세부업종별 추정은 편의가 발생하거나 효율이 떨어지는 추정값을 얻을 가능성이 높다.

따라서 2006년도에 시행된 광업·제조업 조사를 통한 모집단을 구성하여 94개 세부업종별 분석을 실시한다. 세부업종별 분포를 파악하여 표본규모의 산정과 배분에 활용하면 세부업종별로 대표성 있는 표본을 얻을 수 있다. 세부업종별로 생산액, 출하액, 재고액 등을 분석하여 표본설계를 실시하고 이들 정보는 조사 자료를 이용한 추정에서 가중값 계산에 사용한다.

2.3. 층화

기존의 부품·소재산업 동향조사는 12개 업종별로 층화를 하여 12개 업종 및 94개 세부업종별 생산액 및 출하액 등을 추정해오고 있다. 앞서 언급했듯이 정도가 낮은 추정이 될 가능성이 존재하므로 새로운 표본설계에서는 94개 세부업종별 생산액 및 출하액 등의 추정값의 정도를 높이기 위한 표본설계 반영이 필요하다. 정확하고 신뢰할 수 있는 추정을 위해 층화는 94개 세부업종에 대해 실시한다. 94개 세부업종에 대해 1차 층화를 실시한 뒤 분석을 통해 전수조사층과 표본조사층으로 2차 층화를 실시한다. 1차 층화는 한국표준산업분류상의 구분으로 실시하고 2차 층화는 응용절사법(Modified Cut-off Method)을 통해 층화를 실시한다. 단, 주성분 점수를 이용한 네이만 배정방법은 94개 세부업종에 대한 1차 층화에 대해서만 배정한다.

2.4. 보조변수 선정

2006년 광업·제조업 통계조사에서 사용할 수 있는 변수는 생산액, 총출하량, 총출하액, 내수출하금액, 수출출하금액 등이다. 변수들 간의 상관관계를 통해 가장 적당한 보조변수를 선정하고자 한다. 표 2.2에서 보는 바와 같이 생산액이 총출하액, 수출출하액, 내수출하액과 상관관계가 매우 높으므로 생산액을 일변량 보조변수로 사용하여도 설계 및 추정하는 데 아무런 문제가 없다. 따라서 표본설계는 생산액을 기준으로 실시한다. 단, 주성분 점수를 이용한 다변량 네이만 배정방법에서는 5개의 보조변수를 모두 사용한다.

표 2.2: 보조변수들간의 상관관계

	생산액	총출하량	총출하액	수출출하액	내수출하액
생산액	1.000				
총출하량	0.001	1.000			
총출하액	1.000	0.001	1.000		
수출출하액	0.915	0.002	0.918	1.000	
내수출하액	0.788	0.000	0.785	0.474	1.000

3. 표본설계 제안

3.1. 표본설계 개요

3.1.1. 응용절사법

2005년도에 설계된 기존 표본은 2003년 광업·제조업 통계조사 자료를 이용하여 12개 업종을 층화하였다. 새로운 표본설계는 2006년 광업·제조업 통계조사 자료를 이용하여 94개 세부업종별 사업체를 대상으로 표본설계를 하고자 한다. 과거 자료를 분석해서 주요 특성에 대한 94개 세부업종별 생산액, 출하액, 재고액의 변동계수를 구하고 이를 이용해서 일정 수준의 변동이 유지되도록 표본설계를 한다. 또한 표본배정 시에 전체적으로 추정의 정확성을 높이고 층별로 추정의 정도가 일정 수준을 유지되도록 한다. 표본설계는 층별로 응용절사법을 이용하여 각 층별로 표본크기 결정 및 배정을 실시하고 표본추출은 확률비례 계통추출로 실시한다.

층화추출에서 층간 최적경계의 설정에 대해 Sethi (1963)는 모집단이 정규분포, 감마분포, 베타분포를 따르는 경우에 있어 최적경계점을 표로 제시하였다. Dalenius와 Hodges (1959)는 각 층에서 보조변수의 누적도수제곱근이 같게 되도록 하는 누적도수제곱근방법(cumulative frequency square root)을 제안하였다. 누적도수제곱근방법은 층화변수가 인접한 경계점 사이에서 균일분포에 근사한다는 가정 하에서 유도되었기 때문에 왜도가 큰 기업체조사와 같은 모집단에서는 만족할 만한 경계점을 제공하지 못할 수도 있다. 따라서 실제 조사에서는 일정 규모 이상인 업체는 전수조사하고, 상대적으로 왜도가 줄어든 나머지 표본층의 모집단에 대해 누적도수제곱근방법을 적용하여 경계점을 결정하는 방법이 주로 사용되고 있다. Hidiroglou (1994)는 절사법을 이용한 층화추출에 있어 전수층과 표본층의 경계를 결정할 수 있는 객관적인 절차를 제시하였다. Lavalley와 Hidiroglou (1988)은 Sethi의 방법을 확장하여 응용절사법에서 전수층과 표본층의 경계 및 표본층간의 최적경계점을 결정해주는 알고리즘을 작성하여 누적도수제곱근방법과 비교하여 상대적으로 우수함을 입증하였다. 또한 표본층에서 조사 단위들을 단순임의추출하는 경우에 목표 변동계수를 만족하는데 필요한 표본크기를 최소로 하는 전수층 및 표본층들 간의 최적경계점을 결정하기 위해 다음과 같은 반복적 절차(iterative procedure)를 이용하였다.

본 표본설계에서 응용절사법은 각 1차 층별로 각각 계산되기 때문에 먼저 크기가 N_h 인 하나의 층을 보조변수 y 에 따라 $y_{(1)}, y_{(2)}, \dots, y_{(N_h)} (y_i \leq y_{i+1})$ 과 같이 오름차순으로 정렬한 다음 모집단을 전수층과 표본층의 2개의 층으로 나누어 각 층의 $N_{hk} (k = c, s)$ 개 조사

단위로부터 n_{hk} 개의 단위를 비복원 단순임의추출한다. 여기서 c 는 전수층을, s 는 표본층을 나타낸다. 따라서 h 번째 층의 모총합 $Y_h = \sum_{i=1}^{N_{hc}} y_{hci} + \sum_{j=1}^{N_{hs}} y_{hsj}$ 의 추정량은

$$\hat{Y} = \sum_{i=1}^{N_{hc}} y_{hci} + \frac{N_{hs}}{n_{hs}} \sum_{j=1}^{N_{hs}} y_{hsj} \quad (3.1)$$

이 된다. w 를 표본층에서 모집단 크기와 표본크기의 비율이라고 할 때 목표변동계수 CV를 만족하기 위해 필요한 총 표본크기는 다음과 같다.

$$n_h = N_h + \frac{N_{hs}^2 S_{hs}^2 / w}{CV^2 \hat{Y}^2 + N_{hs}^2 S_{hs}^2}, \quad (3.2)$$

따라서 표본층에 배분되는 표본의 크기는 $n_{hs} = n_h - N_{hc}$ 이 된다. 여기서 S_{hs}^2 는 h 번째 층의 표본층의 모분산을 나타낸다. h 번째 층의 총표본크기 n_h 이 최소화되는 최적경계점 $b_h (y_{(1)} < b_h < y_{(N)})$ 은 식 (3.2)의 n_h 를 b_h 에 대해 편미분한 다음 0으로 놓아 정리하면 다음과 같은 이차방정식을 얻게 된다. 여기서 $\alpha_h, \beta_h, \gamma_h$ 는 이차방정식을 만족하게 하는 임의의 실수이다.

$$\alpha_h b_h^2 + \beta_h b_h + \gamma_h = 0. \quad (3.3)$$

응용절사법은 다음과 같은 알고리즘을 이용해서 식 (3.3)의 이차방정식 해를 구한다.

[단계1] $y_{(1)}, y_{(2)}, \dots, y_{(N_h)} (y_i \leq y_{i+1})$ 과 같이 오름차순으로 정렬한 모집단의 조사단위에서 $b_h = y_{(N)}$ 로 놓는다.

[단계2] 임의의 경계점 $b'_h < y_{(N_h)}$ 을 초기값으로 설정한다.

[단계3] [단계1]에서 설정한 초기경계를 이용해서 표본배정비율 N_{hs}/N_h 와 층분산 S_{hs}^2 를 계산한 다음에 초기 경계값을 b''_h 로 다음과 같이 대체한다.

$$b''_h = \frac{-\beta'_h + \sqrt{\beta'^2_h - 4\alpha'_h \gamma'_h}}{2\alpha'_h}. \quad (3.4)$$

[단계4] $\max\{b''_h - b'_h\} < \varepsilon, \varepsilon > 0$ 이 만족할 때까지 [단계3]을 반복함으로써 최적의 경계점을 산출한다.

3.1.2. 주성분점수를 이용한 다변량 네이만 배정

p 개의 다변량 보조변수를 이용하여 네이만 배정을 하고자 한다면 변수들 간의 상관관계가 다양하게 존재한다. 따라서 상관관계를 무시할 수 있도록 주성분분석을 실시하여 p 개의 주성분 점수를 만들고 이를 다변량 보조변수로 간주하고 배정하는 방법이다 (Namkung 등, 2007). 주성분 점수는 완전하게 상관관계가 존재하지 않기 때문에 상관관계를 고려하지 않고 일괄적인 다변량 네이만 배정을 이용할 수 있다.

p 개의 보조변수가 존재하기 때문에 일변량 네이만 배정처럼 분산 상대비율을 이용할 수가 없다. 다변량 네이만 배정방법은 p 개의 보조변수의 분산합의 상대비율을 통해 네이

표 3.1: 12개 업종별 표본배분

업종	조사 모집단	응용절사법				다변량네이만	
		표본크기	전수층	표본층	상대허용오차	표본크기	상대허용오차
합계	21,989	2,073	1,133	940	2.38	2073	2.38
섬유제품	1,388	173	89	84	9.94	54	15.46
화합물 및 화학제품	1,236	175	91	84	6.48	176	3.09
고무 및 플라스틱제품	2,130	194	106	88	9.90	138	6.21
비금속광물제품	449	64	40	24	9.67	75	3.09
제1차 금속제품	820	126	78	48	4.72	216	1.05
조립금속제품	3,391	256	140	116	9.72	160	7.05
일반기계부품	6,124	326	140	186	9.76	178	10.20
컴퓨터 및 사무기기부품	261	43	23	20	9.86	26	7.58
전기기계부품	2,676	236	139	97	9.85	280	3.41
전자부품	1,231	164	107	57	4.33	485	0.51
정밀기계부품	636	84	48	36	9.91	41	8.70
수송기계부품	1,647	232	132	100	8.60	244	3.52

만 배정을 실시하는 방법이다. 주성분 점수는 상관관계가 없고 표준화된 값으므로 분산합이 일변량 분산과 동일한 개념을 갖게 된다. 다만 주성분 점수는 각가 변동을 설명하는 비율이 다르므로 주성분 성분의 고유값을 가중값으로 이용하는 분산 가중합을 이용한다. 이를 적용하면 각 층의 표본크기는 다음과 같다.

$$n_h = n \left[\frac{N_h \sqrt{\text{tr}_w(\Sigma_h)}}{\sum_{h=1}^L N_h \sqrt{\text{tr}_w(\Sigma_h)}} \right], \quad (3.5)$$

여기서 가중분산합은 $\text{tr}_w(\Sigma_h) = \sum_{j=1}^p w_j s_{jh}^2$ 이고 s_{jh}^2 는 h 층의 j 번째 주성분의 분산이다. 가중값인 w_j 는 λ_j/p 로 계산한다.

3.2. 표본배분

전체 허용오차가 최소가 되도록 94개 각 세부업종별 최대허용오차와 표본규모를 조정하고 12개 업종별로 허용오차가 10%가 넘는 업종에 대해서는 10% 미만이 되도록 알고리즘을 확장하여 반복한다. 94개 세부업종별 분산추정이 가능하도록 최소표본크기를 2개 업체로 우선배정한다. 94개 세부업종에 대해 응용절사법으로 배분한 결과 일부 세부업종에서는 최대허용오차가 80%로 나타난 업종이 있었다. 이러한 결과는 세부업종의 모집단 크기가 작거나 세부업종의 비중은 작지만 세부업종 내에 특정 소수의 업체 비중이 매우 커서 적은 수의 표본이 배정된 결과이다.

배정된 표본에 대해 각 층별로 계통추출을 이용하였다. 분산추정을 위해 최소표본크기 2개 업체를 기준으로 추출하였고 94개 업종별로 생산액이 큰 순서로 정렬하여 추출하였다. 계통추출에서 필요한 출발점은 임의로 '추출간격 $\times 0.5$ '로 정하여 추출을 하였으며, 상대허용오차는 비교를 위해 생산액으로 계산하였다.

표 3.2: 94개 세부업종별 표본배분

업종	세부업종	조사 모집단	응용절사법				다변량네이만	
			표본크기	전수총	표본총	상대허용오차	표본크기	상대허용오차
섬유제품	제사 및 방직	289	44	25	19	21.28	11	16.71
	직물제조	599	72	33	39	17.02	13	11.80
	섬유염색 및 가공	169	22	14	8	23.88	8	19.01
	기타 섬유제품	331	35	17	18	18.76	22	16.56
화학물 및 화학제품	석유화학계 기초 유기화학물	4	4	2	2	28.07	3	32.84
	기타 기초 유기 유도체 및 화합물	36	20	10	10	10.25	26	5.05
	기타 기초 무기 화합물	88	16	7	9	36.33	12	14.36
	무기안료, 염료, 유연제 및 기타착색제	146	11	4	7	65.36	5	32.81
	합성고무 및 플라스틱물질	53	27	22	5	7.67	19	5.08
	기초의약품 및 생물학적 제제	38	5	2	3	76.64	4	40.73
	농약	16	5	2	3	54.86	3	36.47
	도료 및 인쇄잉크	185	25	14	11	32.39	10	18.46
	계면활성제	77	7	3	4	74.46	4	34.54
	사진용화학제품 및 감광재료	39	8	5	3	72.07	4	42.33
	그외 기타 화학제품	523	34	13	21	31.00	79	3.95
	화학섬유	31	13	7	6	22.73	7	14.21
고무 및 플라스틱 제품	고무타이어 및 튜브	8	7	5	2	1.08	9	15.66
	제1차플라스틱제품	877	73	38	35	18.76	47	5.76
	기계장비 조립용 플라스틱제품	1,082	102	57	45	16.78	73	2.90
	기타 고무제품	163	12	6	6	51.92	9	23.97
비금속 광물제품	판유리	3	3	3	0	0.00	3	20.81
	산업용 유리제품	200	34	26	8	10.31	60	7.92
	산업용도자기	39	5	3	2	69.17	3	67.66
	구조용 정형내화제품	6	4	2	2	38.26	2	59.91
	석회	10	2	0	2	51.80	2	24.18
제1차 금속제품	기타 비금속광물제품	191	16	6	10	47.80	5	31.03
	합금철	8	3	1	2	42.27	3	43.39
	철강압연, 압출 및 연신제품	60	20	15	5	4.10	71	3.83
	철강관	319	36	20	16	23.10	75	3.28
	비철금속 재련, 정련 및 합금	66	15	10	5	13.07	20	11.06
	비철금속압연, 압출 및 연신제품	220	36	22	14	17.94	34	6.11
	철강주조업	133	8	4	4	54.65	6	30.60
조립 금속제품	철강주조업(절단가공)	14	8	6	2	11.12	7	17.29
	금속탱크, 저장조 및 유사용기	250	26	17	9	23.31	26	19.41
	원자로 및 증기발생기부품	36	5	3	2	65.75	2	89.98
	분말아금, 금속단조, 압형제품	1,030	100	53	47	15.89	66	3.55
	금속열처리업, 도금 및 기타처리업	1,109	45	17	28	27.59	16	12.41
	톱 및 호환성공구	236	17	12	5	35.18	8	39.56
	금속파스너 및 스프링	644	51	30	21	22.81	35	6.07
	기타 금속제 부품	86	12	8	4	32.87	7	37.42
	내연기관 및 터빈	108	11	7	4	15.37	16	18.18
	펌프 및 압축기	445	33	21	12	30.20	26	12.06
일반 기계부품	탱, 밸브 및 유사장치	453	26	8	18	42.53	10	14.97
	베어링, 기어 및 동력전달장치	318	22	13	9	35.77	18	14.71
	산업용오븐 및 노용부품	41	5	3	2	69.48	2	71.43
	운반하역기계부품	235	9	2	7	70.43	4	28.47
	냉동공조기계부품	869	65	29	36	26.07	27	6.88
	가공공작기계부품	255	7	0	7	78.59	4	29.07
	농업용기계부품	167	6	1	5	78.09	4	76.47
	금속주조 및 기타 야금용 기계 부품	26	4	2	2	79.28	3	70.90
	건설광산기계부품	180	16	7	9	45.31	10	25.54
	음식료품기계 및 담배가공기계부품	15	3	1	2	68.73	2	39.93
	섬유기계 부품	152	5	0	5	72.72	2	45.14

업종	세부업종	조사 모집단	응용절사법				다변량네이만	
			표본크기	전수층	표본층	상대허용오차	표본크기	상대허용오차
	반도체제조장비부품	219	15	8	7	56.79	7	24.40
	무기 및 총포탄 부품	22	8	6	2	18.73	9	13.73
	가정용기기부품	310	19	5	14	47.59	9	22.71
	금형	1,819	59	23	36	28.28	20	8.02
	기타 일반기계부품	490	13	4	9	60.44	5	24.04
컴퓨터 및 사무기기 부품	컴퓨터 기억장치	4	4	2	2	5.27	2	13.42
	컴퓨터 입력장치	51	6	2	4	71.93	11	51.53
	컴퓨터 단말기	31	5	2	3	69.81	2	48.45
	컴퓨터용 메인보드 및 카드	18	5	3	2	54.22	2	52.20
	컴퓨터 주변기기 및 전용부품	118	18	11	7	31.81	6	30.40
	사무용기기부품	39	5	3	2	59.67	3	70.53
전기 기계부품	전동기, 발전기, 전기변환장치	692	67	41	26	20.86	46	5.60
	전기공급 및 전기제어장치	1,338	73	38	35	19.76	136	3.41
	절연선 및 케이블	284	34	20	14	21.58	12	12.95
	절연광섬유케이블	6	4	2	2	34.25	3	45.60
	일차전지 및 축전지	26	9	7	2	22.87	13	14.38
	전구 및 램프부품	26	4	2	2	37.51	4	60.80
	조명장치 부품	65	6	4	2	64.42	3	51.18
기타 전기장비 부품	239	39	25	14	24.20	63	2.90	
전자부품	개별소재 반도체	94	14	10	4	34.63	8	20.91
	집적회로 반도체	27	11	7	4	5.11	17	6.28
	통신기기 및 방송장비	577	44	24	20	28.17	41	7.13
	전자관	5	4	2	2	22.69	3	31.18
	인쇄회로판	169	34	26	8	21.90	114	1.39
	전자축전기	38	7	5	2	57.58	6	31.69
	전자저항기	31	3	1	2	69.97	7	48.05
	전자카드	29	4	1	3	77.35	3	68.20
	액정표시장치	45	25	18	7	3.84	267	0.93
	그외 기타 전자부품	216	18	13	5	38.36	19	17.32
정밀 기계부품	방사선장치 및 전기식 진단기기부품	17	4	2	2	63.33	2	55.91
	기타의료용기기 및 부품	65	7	4	3	74.69	3	63.32
	측정,시험,항해,기타정밀기기 및 부품	477	54	27	27	14.67	121	15.44
	광섬유 및 광학요소	53	16	14	2	12.10	13	24.26
	시계부품	24	3	1	2	66.92	2	70.02
수송 기계부품	자동차용 엔진	5	3	1	2	24.83	5	9.51
	자동차부품	1,390	199	114	85	9.32	217	1.49
	선반구성부분품	130	10	5	5	56.10	8	25.48
	철도차량부품 및 관련장치	58	7	5	2	65.12	4	56.38
	항공기용 엔진 및 부품	10	5	3	2	35.23	6	27.85
	모터사이클 부품	44	5	3	2	74.81	2	70.93
	자전거부품	10	3	1	2	60.76	2	38.15
	합계	21,989	2,073	1,133	940		2,073	

상대허용오차를 계산한 결과 응용절사법은 예상 상대허용오차가 고른 반면 다변량 네이만 배정은 섬유제품과 일반기계부품의 예상 상대허용오차가 큰 것으로 나타났다. 이는 두 업종의 분산이 크기 때문인데 응용절사법은 전수층과 표본층으로 나누어 분산이 줄어드는 효과가 있기 때문에 예상 상대허용오차를 낮추는 효과가 있다.

응용절사법은 5개의 조사지표 중 생산액을 이용하여 표본배분을 실시하였기 때문에 생산액 뿐만 아니라 생산액과 양의 높은 상관관계를 보이는 총출하액, 수출출하액, 내수출하액에 대한 표본오차를 줄이는데 좋은 방법이다. 그러나 생산액과 상관관계가 높지 않은 총

출하액의 경우는 표본오차를 줄이는데 효과적이지 않는 단점을 가지고 있다. 반면 주성분 점수를 이용한 다변량 네이만 배정의 경우는 5개의 조사지표를 동일한 가중에 의해 동시에 고려하여 표본배분을 실시하였기 때문에 표본오차를 동시에 줄이는데 효과적이다.

4. 추정

4.1. 가중값

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

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

먼저 표본추출 과정을 살펴보면 모집단을 94개 층으로 층화한 후에 각 층에서 다시 전수층과 표본층으로 층화하여 사업체를 추출했기 때문에 2단 층화추출이다. 따라서 추출가중값은 이를 모두 적용해야 한다. 또한 앞서 분석한 대로 층별로 생산액의 변동이 매우 크기 때문에 이에 따른 변동 역시 보정해 주어야 한다. 표본조사층의 세번째 가중값을 제외하고는 모두 각 층에서 동일한 가중값을 가진다. 특히 전수조사층은 모두 가중값 1을 가진다. 가중값 계산을 위한 기호는 다음과 같다.

W_{hcj} : h 층의 전수조사층의 j 번째 사업체의 가중값.

W_{hsi} : h 층의 표본조사층의 i 번째 사업체의 가중값.

N_{hc} : h 층의 전수조사층의 모집단 크기.

N_{hs} : h 층의 표본조사층의 모집단 크기.

n_{hs} : h 층의 표본조사층에서 추출된 사업체 표본크기.

R_{hc} : h 층의 전수조사층에서 조사된 사업체 표본크기.

R_{hs} : h 층의 본조사층에서 조사된 사업체 표본크기.

Q_{hs} : h 층의 표본조사층의 모집단 사업체의 생산액 총합.

Q_s : 모집단 사업체의 생산액 총합.

1. 설계가중값(1): 추출확률만을 이용하여 각 층 사업체 수로 보정한 가중값

$$W_{hcj}^{(1)} = 1, \quad W_{hsi}^{(1)} = \frac{N_{hs}}{n_{hs}}. \quad (4.1)$$

2. 가중값(2): 설계가중값에 항목 무응답을 보정한 가중값

$$W_{hcj}^{(2)} = \frac{N_{hc}}{R_{hc}}, \quad W_{hsi}^{(2)} = \frac{N_{hs}}{R_{hs}}. \quad (4.2)$$

3. 가중값(3): 실제가중값에 항목 무응답을 보정하고 각 94개 세부 업종의 생산액을 고려한 가중값

$$W_{hcj}^{(3)} = \frac{N_{hc}}{R_{hc}}, \quad W_{hsi}^{(3)} = \frac{N_{hs}Q_{hs}}{R_{hs}Q_s} \cdot \frac{N}{\sum_{h=1}^H \sum_{i=1}^{n_{hs}} \frac{N_{hs}Q_{hs}}{R_{hs}Q_s}}. \quad (4.3)$$

최종 표본단위 가중값은 세 번째 가중값을 사용하게 된다. 전수조사층의 경우 전수조사하기 때문에 생산액을 고려하지 않으므로 세 번째 가중값이 두 번째 가중값과 동일하다. 다변량 네이만 배정의 경우는 모든 가중값을 표본층으로 가중하고 계산한다.

4.2. 특성별 추정

94개 세부업종별로 전수조사층과 표본조사층으로 층화하였기 때문에 모평균 추정은 의미가 없다. 전수조사층에서 모두 조사하였고 표본조사층은 일부만 조사되었기 때문에 모평균 추정값은 과대추정되기 때문이다. 따라서 94개 세부업종별, 12개 업종별, 전체 모층합 추정을 해야 한다.

먼저 94개 세부업종별 층합추정은 다음과 같이 계산된다.

$$\hat{\tau}_h = \sum_{j=1}^{N_{hc}} W_{hcj} y_{hcj} + \sum_{i=1}^{n_{hs}} W_{hsi} y_{hsi}, \quad (4.4)$$

여기서 $\hat{\tau}_h$ 는 h 층(h 번째 세부업종)의 층합추정량, y_{hcj} 는 h 층의 전수조사층의 j 번째 관측값이고 y_{hsi} 는 h 층의 표본조사층의 i 번째 관측값이다. 다변량 네이만 배정의 경우는 모든 사업체를 표본층으로 가정하고 계산한다. 즉, 전수층 층합추정이 0이므로 층합추정은 두 번째 항만을 이용하여 계산한다.

94개 세부업종별 층합추정에 대한 분산추정량은 표본조사층에서만 구한다. 전수조사에서는 표본오차가 없기 때문이다.

$$\hat{V}(\hat{\tau}_h) = N_{hs}^2 \left(\frac{N_{hr} - n_{hs}}{N_{hs}} \right) \left(\frac{n_{hs}}{n_{hs} - 1} \right) \sum_{i=1}^{n_{hs}} (e_{hsi} - \bar{e}_{hs})^2, \quad (4.5)$$

여기서 $e_{hsi} = \{W_{hsi}(y_{hsi} - \bar{y}_s)\}/N_{hs}$, $\bar{e}_h = \sum_{i=1}^{n_{hs}} e_{hsi}/n_{hs}$ 이다. 그리고 94개 세부업종별 변동계수의 추정량은 $CV = \sqrt{\hat{V}(\hat{\tau}_h)}/\hat{\tau}_h$ 이다.

12개 업종에 대한 추정은 일반적인 층화추출의 공식을 이용하면 된다. 본 연구에서는 층을 94개로 나누었기 때문에 12개 업종 구분에 따른 기호를 사용하지 않았다. 예를 들어 94개 세부업종 중 $h=1, 2, 3, 4$ 가 12개 업종 중 A 업종을 의미한다면 모층합 추정량과 분산추정량을 구할 수 있다.

$$\hat{\tau}_A = \sum_{h=1}^4 \hat{\tau}_h, \quad \hat{V}(\hat{\tau}_A) = \sum_{h=1}^4 \hat{V}(\hat{\tau}_h). \quad (4.6)$$

따라서 이를 확장하면 부품·소재산업 전체 특성값의 모총합 추정량과 분산추정량을 얻을 수 있다.

$$\hat{\tau} = \sum_{h=1}^{94} \hat{\tau}_h, \quad \hat{V}(\hat{\tau}) = \sum_{h=1}^{94} \hat{V}(\hat{\tau}_h). \quad (4.7)$$

이에 따라 전체 업종의 변동계수 추정량은 $CV = \sqrt{\hat{V}(\hat{\tau})/\hat{\tau}}$ 이다.

5. 결론

“부품·소재 동향조사”는 계속 반복되는 조사이므로 전반적인 시계열 유지가 매우 중요하다. 그러나 사업체 모집단과 조사대상 및 표본설계가 빠른 속도로 변화함에 따라 전체적으로 시계열 유지에 어려움이 있다.

2005년도에 설계된 표본에서는 12개 업종별로 층화를 하여 94개 세부업종에 따른 추정치는 편이가 발생하거나 효율이 떨어질 수 있는 가능성을 내포하고 있었다. 12개 업종별 설계로 우리나라 부품·소재산업 현황을 파악하기보다는 94개 세부업종으로 층을 좀 더 세분화하면 현황을 파악하는데 더 좋은 추정값을 얻을 수 있다.

제한한 표본설계의 특징은 가장 최근 자료인 2006년 통계청이 시행한 광업·제조업 통계조사 결과를 모집단으로 한 표본설계를 통해 현재의 상황을 반영한 산업에 관한 정보 제공이 가능한 점이다. 보다 세분화된 업종에 따라 매월 부품·소재산업의 생산, 출하, 재고의 변동사항을 조사하여 부품·소재산업의 경기변동실태를 파악하고 부품·소재산업의 육성 정책 및 기업경영의 기초자료를 제공할 수 있도록 표본설계를 실시하였다.

“부품·소재 동향조사”는 생산액, 총출하량, 총출하액, 수출출하액, 내수출하액 등 5개의 조사지표에 의해 실시된다. 응용절사법은 5개의 조사지표 중 생산액을 이용하여 표본배분을 실시하였고 주성분 점수를 이용한 다변량 네이만 배정의 경우는 5개의 조사지표를 동일한 가중치에 의해 동시에 고려하여 표본배분을 실시하였다. 따라서 생산액에 가중치를 주어 분석한다면 응용절사법이 효과적일 것이고 생산액과 양의 높은 상관관계를 보이는 총출하액, 수출출하액, 내수출하액에 대한 표본오차도 줄어드는 결과는 얻을 것이다. 그러나 생산액과 상관관계가 높지 않은 총출하액의 경우는 표본오차를 줄이는데 효과적이지 않기 때문에 5개 조사지표에 대해 모두 동일하게 중요성을 두고 분석할 경우는 주성분 점수를 이용한 다변량 네이만 배정방법이 표본오차를 줄이는데 더 효과적일 것이다.

앞으로도 변화된 모집단에 적합하고 통계조사 결과의 비교성과 대표성을 높이기 위해 표본설계에 대한 연구가 지속적으로 이루어져야 하며, 부품·소재산업 동향조사가 표본의 대표성을 충분히 높일 수 있을 때까지는 매년 또는 2년 주기로 새로운 표본설계를 통해 모집단 및 추출틀을 교체하여 조사의 전반적인 내용들이 컨설팅될 수 있도록 하는 것이 바람직할 것으로 생각된다. 이처럼 표본설계에 대하여 체계적이고 지속적인 연구가 있을 때 질 높은 부품·소재산업의 동향 통계 생산이 가능할 것이다.

한편, 통계분석 과정에서 가중값을 무시하고 분석한 추정값은 심각한 편이가 발생할 수 있다. 표본의 크기가 큰 대규모 조사에서 문제가 되는 것은 표본오차 내지 허용오차가 아

나라 추정량의 편의이기 때문에 추정과정에서 반드시 가중값을 이용해야 한다. 부품·소재 산업 동향조사와 같은 복합표본조사의 가중값은 설계 가중값, 무응답에 대한 조정, 사후층화 또는 보조변수에 대한 조정 등의 세 가지 요인을 통합하여 산정되어야 하며, 결과를 발표할 때 이러한 내용들을 고려하였는지를 반드시 검토하여야 하고, 이에 대한 연구가 필요하다.

최근 들어 조사환경이 열악해져서 조사대상업체로부터 조사 협조를 얻기 어려운 상황에 놓여 응답률이 떨어지거나 표본대체율이 증가하는 등 비표본오차의 발생 가능성이 증가하고 있다. 특히, 사업체 대상 표본조사에서는 무응답 사업체의 경우 표본대체를 통해 응답을 얻게 되는 데, 이때의 표본대체율도 관리가 되어야 한다. 이러한 어려운 상황을 극복해야만 효율적인 부품·소재산업 현황 통계생산이 가능해질 것으로 보인다.

아울러 사업체 규모 변동에 대처하는 방안, 부품·소재 사업체의 창업, 휴·폐업, 전업 등의 처리 방안, 무응답 대처 방안 등 사업체 조사 과정에서 나타나는 비표본오차를 줄이기 위한 좀 더 심층적인 연구가 필요하다.

참고문헌

- 한국기계산업진흥회 (2007). <부품·소재 동향보고서>, 대한통계협회.
- Dalenius, T. and Hodges, J. L. (1959). The choice of stratification points, *Skandinavisk Aktuarietidskrift*, **40**, 198-203.
- Hidiroglou, M. (1994). Sampling and estimation for establishment surveys: Stumbling blocks and progress, In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 153-162.
- Lavallee, P. and Hidiroglou, M. (1988). On the stratification of skewed populations, *Survey Methodology*, **14**, 33-43.
- Namkung, P., Byun, J. and Choi, J. (2007). Multivariate Neyman allocation using principle component analysis, *ISL*, 438.
- Sethi, V. K. (1963). A note on optimum stratification of populations for estimating the population means, *Australian Journal of Statistics*, **5**, 20-33.

[2008년 8월 접수, 2008년 9월 채택]

Sample Design for Materials and Components Industry Trend Survey[†]

Pyong Namkung¹⁾

Abstract

This paper provides correct informations inflecting the present situation using the sample design in population that the National Statistical Office puts in operation of the mining and manufacturing industry statistical survey in 2006. This paper proposes new sampling design which is able to grasp business fluctuations and provide basic data for the rearing policy and management of the material industry and components industry. These sample design are the modified cut-off method and multivariate Neyman allocation using principal components and sampling method is the probability proportional systematic sampling

Keywords: Sample design; sampling error; modified cut-off method; PPS systematic sampling.

[†] This research was supported by Sungkyunkwan University Research Grant, 2007.

¹⁾ Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.
E-mail: namkung@skku.edu