

시군구 실업자 추정을 위한 소지역 추정법

이계오¹⁾

요약

지방자치제의 정착으로 시군구 단위의 통계 작성이 요구되었을 뿐만 아니라 IMF사태로 실업자 통계가 정치·사회적인 이슈가 되면서 시군구별 실업자통계 작성의 필요성이 제기 되었다. 현재의 경찰조사 체계에서 시군구별 실업통계를 작성할 수 있는 2종의 소지역 추정법을 제안하고 실제 경찰조사 자료를 이용하여 충북 시군구별 실업자를 추정하고 추정법의 효율을 비교하여 소지역 추정법의 적용 가능성을 입증하였다.

주요용어: 소지역 추정법, 합성 추정량, 복합 추정량.

1. 서론

우리나라에서는 취업, 실업 등과 같은 경제적 특성을 조사하여 국가의 고용정책 입안과 평가에 필요한 기초자료를 수집하기 위해서 매월 3만 표본가구내에 거주하는 만 15세 이상인 사람들을 대상으로 경제활동 전반에 걸쳐서 조사하고 있다. 매월 15일을 조사기준일로 정하고 15일을 포함하는 조사대상 주간중에 표본가구에 거주하는 사람들의 취업, 실업 및 비경제활동인구 관련사항을 방문면접이나 컴퓨터 면접방식으로 조사한 후 조사된 자료를 이용하여 매 익월 말경에 전국단위와 각 시도단위로 경제활동 관련 통계를 발표하고 있다.

1995년부터 시작된 지방자치제도와 1997년의 IMF사태로 실업자 구제대책의 효과적인 시행을 위해서 각 시도내의 시군구별 실업관련 통계작성이 요구되었을 뿐만 아니라 지방자치 행정의 발전적인 정착을 위해서는 다양한 분야의 통계도 시군구 단위까지 작성해야 한다는 인식이 높아지고 있지만 현행의 통계작성 방법으로 모든 분야의 통계를 시군구 단위까지 작성하기 위해서는 현재의 몇배의 시간, 예산과 인원이 소요될 것이므로 실현 가능성이 희박하다. 그러나 통계선진국인 미국이나 캐나다에서 소지역의 통계작성에 이용되고 있는 소지역 추정법을 우리나라에서도 적용할 수 있다면 현재의 예산과 인원규모로도 시군구 단위의 통계를 작성할 수 있음을 살펴보고자 한다.

본 연구는 경제활동인구조사에서 실업자 수 추정만으로 제한하여 현재 활용되고 있는 표본설계를 이용하여 시군구 실업자 수를 추정할 수 있는 방법에 중점을 두고 있다. 2절에서는 소지역 추정법을 이용한 2가지 추정량과 이들의 비교를 위한 직접추정량을 설명하고, 3절에서는 경제활동인구조사의 실제자료를 이용하여 충청북도 시군구의 실업자 수 추정값과 추정오차를 계산하고 추정량의 효율을 비교하며, 마지막으로 연구내용의 실제 적용성과 앞으로 해결되어야 할 과제를 요약하였다.

1) (363-849) 충북 청원군 남일면 쌍수리, 공군사관학교 전산통계학과, 교수
E-mail: kayolee@hammail.com

2. 소지역 추정법

표본설계 당시에는 통계작성단위의 규모가 컸으나 자료 수집 후에 통계작성단위의 세부단위에 대한 통계를 생산할 필요가 있을 경우에는 세부단위에서 조사된 자료의 수가 적거나 아예 없을 수 있으므로 세부단위에 대한 추정값은 추정오차가 클 뿐만 아니라 세부단위들간의 표본수의 차이 때문에 추정오차의 크기도 천차만별이 될 것이다. 예를 들면 경찰조사에서 시도별로 실업 취업 등의 통계를 생산하고자 한다면 각 시군구에 할당된 표본조사구 수의 분포가 불균등하므로 시군구 단위별 실업통계의 신뢰성이 크게 저하될 수 있으며 또 다른 경우로는 각 시도에서 인구사회학적 분류인 성별-연령대별로 구분하여 각 Cell에서 실업률이나 실업자에 대한 통계를 구하고자 할 경우에도 각 Cell에 분포된 표본수가 불균형적일 뿐만 아니라 적기 때문에 추정오차가 커질 것이다.

세부단위에 배정된 표본수가 적을 경우에는 특성이 유사한 인근의 세부단위들을 결합하여 세부단위들의 그룹을 만들고 그 그룹 내에서 세부단위들이 동일한 특성을 갖는다는 조건에서 조사된 자료를 이용하여 세부단위의 통계를 작성하거나, 행정업무 자료 또는 센서스 등 다른 통계조사 정보를 이용하여 세부단위의 통계를 작성하는 기법을 소지역 추정법(Small Area Estimation)이라고 한다.

다음에는 경찰조사에서 소지역 추정법의 적용 가능성을 알아보기 위해서 시군구 단위에 실업자 수 추정을 위한 방법으로 합성추정법과 복합 추정법을 주로 다루며 이 두 추정법의 효율성을 알아보기 위해서 먼저 경찰조사에서 시도단위 통계작성에 사용되는 직접 추정량을 요약하였다.

본 연구에서는 용어의 편의성을 위하여 앞으로는 세부단위를 시군구로 한정하겠으며 연구변수도 실업자 수로 한정하겠다.

2.1. 직접 추정법(DIRECT ESTIMATION)

경찰조사에서 해당 시군구에 배정되어 조사된 조사구들만을 이용하여 해당 시군구의 실업자 수를 추정하는 형식이다.

i 시군구의 실업자 수에 대한 추정량은 조사 모집단과 표본간의 관계에서 산출한 승수(ξ)와 관찰된 자료(y)들의 일차결합으로 표현할 수 있다.

$$\hat{Y}_i = \sum_{j \in s_i} \xi_j y_j \quad (2.1)$$

여기서 s_i 는 i 시군구에서 조사한 조사구들의 집합이다.

식 (2.1)에서 주어진 추정량은 불편추정량이 되도록 ξ_j 를 산정 하였으나 해당 시도내의 시군구별 조사구 수의 불균형적 분포로 추정량의 변동도 불균형적 분포를 갖기 때문에 추정량의 신뢰성 조정에 문제가 있을 수 있다.

2.2. 합성 추정법(SYNTHETIC ESTIMATION)

추정하고자 하는 소지역과 특성이 유사한 다른 소지역들의 정보를 이용하여 추정값의 정도를 높이고자하는 추정방식을 합성 추정법이라 하며, 주변이나 유사지역의 정보를 이용하므로 “Borrow Strength”라고 말하기도 한다.

표본조사의 설계 시에는 대영역에 대해서만 직접 추정값을 구하고자 하였으나 대영역을 분할한 소지역의 추정값이 필요할 때는 대영역과 소지역의 구조적 특성이 같다는 조건 하에서 소지역의 연구변수에 대한 추정값을 구할 수 있는데 이때 대영역의 분할은 지리적인 분할보다는 특성에 따른 분할을 뜻하며 예를 들어서 연령대별로 분할하거나 교육정도별로 분할하는 것을 말한다. 특성의 구조에 대한 구성비가 대영역이나 소지역에서 유사할 것으로 가정하는 것은 실제 우리나라의 경찰조사에서 동부와 읍면부로 나누어서 광역단체별로 층화추출하였으므로 시 지역과 읍면 지역으로 그룹을 나눈다면 각 그룹 내에서는 연령대별 구조나 교육정도별 구조의 특성이 거의 유사할 것으로 판단되므로 시군구 실업자 추정에서 합성 추정법의 이용도 가능할 것으로 생각된다.

대영역을 I 개 소지역으로 분할하며 또한 특성 기준에 따라 j 개 범주로 분류한다면 i 소지역의 추정값은 아래 식으로 구할 수 있다.

$$\hat{Y}_i = \sum_j^J p_{ij} \hat{Y}_j \tag{2.2}$$

단, p_{ij} 는 소지역 범주의 가중값이며 이는 센서스나 행정자료에서 구해질 수 있다.

\hat{Y}_j 는 대영역에서 j 범주에 대한 표본에서 구한 추정값이다. 단 대영역의 표본의 수는 충분하게 많아서 신뢰성 있는 추정값을 구할 수 있다고 가정한다. 시·군의 실업자 추정에 관한 경우를 생각해 보자.

- Y_{ij} : i 시군의 j 범주(연령대별 또는 교육정도별)의 실업자수
- X_{ij} : i 시군의 j 범주(연령대별 또는 교육정도별)의 경제활동인구
- $\hat{Y}_j = \sum_i Y_{ij}$ (j 범주의 대영역 (충청북도의 시지역 또는 군지역)에 대한 합계)이고
- $Y_i = \sum_j Y_{ij}$ (i 시군의 실업자수)이다.

\hat{Y}_j 의 직접 추정값 $\hat{Y}_{d,j}$ 는 표본조사 자료만으로 추정가능하고, X_{ij} 는 센서스 또는 행정자료 등 보조변수의 정보에서 계산 가능한 것으로 가정한다면 합성 추정법은 아래와 같이 나타낼 수 있다.

$$\hat{Y}_i^s = \sum_j \left(\frac{X_{ij}}{X_j} \right) \hat{Y}_{d,j} \tag{2.3}$$

만일에 $\hat{Y}_{d,j}$ 가 비 추정량의 형식을 갖는다면,

$$\hat{Y}_{d,j} = \left(\frac{\hat{Y}_j}{\hat{X}_j} \right) X_j$$

로 나타낼 수 있으므로

$$\hat{Y}_i^s = \sum_j X_{ij} \left(\frac{\hat{Y}_j}{\hat{X}_j} \right) = \sum_j \left(\frac{X_{ij}}{\hat{X}_j} \right) \hat{Y}_j$$

가 된다. 여기서 \hat{Y}_i^s 가 불편추정량이 되기 위해서는 $\frac{Y_j}{X_j} = \frac{Y_i}{X_i}$ 를 만족해야 하고, 이를 만족하지 못할 경우에는 편향추정량이 되고, 이때 \hat{Y}_i^s 의 편향(Bias)은 $B(\hat{Y}_i^s) = E(\hat{Y}_i^s - Y_i)$ 이다. 즉, $B(\hat{Y}_i^s) = \sum_j X_{ij} \left(\frac{Y_j}{X_j} - \frac{Y_i}{X_i} \right)$

\hat{Y}_i^s 의 평균제곱오차 ($MSE(\hat{Y}_i^s)$) 의 근사적 불편추정량이 아래와 같이 주어질 수 있다.

$$\widehat{MSE}(\hat{Y}_i^s) = (\hat{Y}_i^s - \hat{Y}_i)^2 - \widehat{Var}(\hat{Y}_i) \quad (2.4)$$

2.3. 복합 추정법 (COMPOSITE ESTIMATION)

소지역에 배정된 표본수가 적기 때문에 표본 조사만을 이용한 직접 추정량의 불안정에서 오는 낮은 신뢰성과 합성추정량의 편향을 보완하기 위해서 직접 추정값과 합성 추정값의 가중평균은 사용하는데 이를 복합 추정량(Composite Estimator)이라 한다.

$$\hat{Y}_i^c = w_i \hat{Y}_i + (1 - w_i) \hat{Y}_i^s \quad (2.5)$$

여기서 \hat{Y}_i 는 표본조사에서 직접 계산한 추정값이며, \hat{Y}_i^s 는 합성 추정값을 나타낸다. w_i 는 가중값으로 0과 1 사이의 값이다.

먼저 평균제곱오차 $MSE(\hat{Y}_i^c)$ 를 최소화하는 w_i 는 아래와 같다.

$$w_{i(opt)} = \frac{MSE(\hat{Y}_i^s)}{MSE(\hat{Y}_i^s) + V(\hat{Y}_i)} \quad (2.6)$$

최적 가중값 $w_{i(opt)}$ 의 추정값은 다음 식으로 계산된다.

$$\hat{w}_{i(opt)} = \frac{mse(\hat{Y}_i^s)}{(\hat{Y}_i^s - \hat{Y}_i)^2}$$

다음에는 모든 소지역에 공통 가중값을 부여하는 방법으로 초기 공통 가중값 w 를 이용해서 $MSE(\hat{Y}_i^s)$ 들의 평균을 최소화하는 가중값은 아래와 같다.

$$\hat{w}_{i(opt)} = 1 - \frac{\sum \hat{V}(\hat{Y}_i)}{\sum_i (\hat{Y}_i^s - \hat{Y}_i)^2} \quad (2.7)$$

마지막으로 각 소지역에 배정된 표본 크기에 의존하는 가중값이 다음과 같이 계산된다.

$$w_i(\delta) = \begin{cases} 1, & \hat{N}_i \geq \delta N_i \\ \frac{\hat{N}_i}{\delta N_i}, & \text{그외} \end{cases}$$

단 N_i 는 i 소지역의 크기이며 $\hat{N}_i = N(n_i/n)$ 이다. \hat{N}_i 는 직접추정량이며 δ 는 합성추정량의 기여도를 조정하는 값이므로 주관적으로 결정한 값이다. 예를 들어 캐나다 노동력 통계조사에서는 2/3으로 한다.

어떤 추정법에 의해서 소지역의 추정값을 구하더라도 대지역을 소지역으로 분할하여 각 소지역의 추정값을 추정하므로 소지역의 추정값의 합계는 대지역의 추정값과 같아야 할 것이다. 왜냐하면 매월 통계청에서 발표하는 광역시와 도의 실업자수와 해당 소지역의 추정값의 합계가 같도록 조정하지 않으면 서로 상이한 통계수치로 인하여 혼란을 줄 수 있으므로 한 가지 통계 수치가 되도록 조정된 추정량을 계산해야 할 것이다. 여기서는 각 소지역의 추정량은 합성 추정법과 복합 추정법 중 어느 한 방법으로 계산한 것으로 간주할 때 조정된 소지역 추정량은 아래와 같다.

$$\hat{Y}_i^A = \left(\frac{\hat{Y}_i^*}{\sum_i \hat{Y}_i^*} \right) \hat{Y} \tag{2.8}$$

단, \hat{Y} 는 광역시·도의 직접 추정값이며, \hat{Y}_i^* 는 소지역의 *추정법으로 추정한 것이다.

3. 충북 시군구의 실업자 추정

3.1. 개요

충북의 행정구역은 2구 2시 8군으로 편성되어 있다. 다음의 표3.1은 1999년 4월의 경제활동인구 조사 결과를 요약한 것이다.

표 3.1: 충북의 경제활동인구 총괄

(단위 : 천명, %)

구분	15세이상인구	경제활동인구	비경제활동인구	경제활동참가율
남	501	361	140	72.06
여	570	287	283	50.35
전체	1,071	648	423	60.50

표3.1에 따르면 충북의 경제활동 참가율이 남자의 경우, 여자에 비해 비교적 높은 것으로 나타나고 있다. 다음 표3.2는 충북지역의 1999년 4월 시와 군 지역을 구분하여 성별에 따른 경제활동 인구와 조사구 수의 현황을 나타낸 것이다

3.2. 시군구 실업자 추정

3.2.1. 직접 추정량

경찰조사 자료를 이용하여 시군구 실업자를 추정할 수 있는 방법으로써 앞 절에서 3가지를 언급하였으며 이 중에서 직접추정량은 사전에 계산한 동부와 읍면부에 대한 승

표 3.2: 시·군 및 군별 경제활동인구와 조사구 수
(단위 : 명, 개)

시·군	남	여	전체	조사구
청주상당구	46,449	38,615	85,064	8
청주홍덕구	77,055	54,506	131,561	14
충주시	53,813	41,745	95,558	11
제천시	41,417	27,383	68,800	4
소계	218,734	162,249	380,983	37
청원군	26,828	22,235	49,063	5
보은군	14,467	13,453	27,920	2
옥천군	22,294	18,431	40,725	3
영동군	22,463	17,353	39,816	3
진천군	12,138	10,197	22,335	1
괴산군	16,244	16,247	32,491	5
음성군	14,895	14,781	29,676	6
단양군	13,030	11,331	24,361	2
소계	142,359	124,028	266,387	27

수(통계청(1996),pp 12)와 경찰조사 자료만을 이용하여 식 (2.1)을 통해서 계산하였다. 또한 분산은 현재 통계청에서 사용하고 있는 연속 차의 분산공식을 적용하여 계산하였으며 시군구별 실업자 추정값과 분산은 표3.3에 주어졌다. 특기사항은 진천군의 표본조사구는 1개이므로 분산을 계산할 수 없어서 일반화 분산 함수(GVF: Generalized Variance Function)(Wolter(1985),pp 210)를 통해서 계산하였다.

3.2.2. 합성 추정량

여기서는 "Borrow Strength"를 적용하기 위해서 충북을 시 지역과 군 지역으로 크게 2개 그룹으로 구분하고 각 그룹내에서 유사성질 범주의 구분을 성별-연령대별과 성별-교육 정도별로 하고 각 셀에 대한 실업률을 추정하였다.

시 지역 내에서는 각 시의 범주별 실업률은 동일하고 군 지역에서 각 군의 범주별 실업률이 동일하다는 조건 하에서 소지역별로 실업자를 다음과 같이 추정한다.

1. 성별-연령대별 구성비를 이용한 시군구별 실업자 추정

연령대별 구분은 15-24세, 25-34세, 35-44세, 45-54세, 55세 이상으로 5개 범주로 나누어서 실업률 추정의 정도를 높이고자 하였다.

표 3.3: 시군구별 실업자 추정값과 분산

구분		직접 추정량	합성 I	합성 II	복합 I	복합 II
		(통계청)	성별 -연령대	성별 -교육정도	직접 +합성 I	직접 +합성 II
청주 상당구	추정값	6,847	4,832	5,220	6,837	6,838
	분산	4,149,369	21,413	22,823	21,303	22,699
청주 홍덕구	추정값	8,864	8,002	8,522	8,853	8,859
	분산	3,200,521	43,016	50,192	42,446	49,417
충주시	추정값	3,234	5,232	5,435	3,265	3,277
	분산	1,758,276	27,833	35,295	27,399	34,600
제천시	추정값	3,186	4,065	2,954	3,177	3,157
	분산	1,181,569	15,758	18,807	15,551	18,513
청원군	추정값	905	1,488	1,370	911	911
	분산	724,201	7,470	9,010	7,394	8,899
진천군	추정값	784	817	790	784	784
	분산	184,041*	2,548	2,156	2,513	2,131
괴산군	추정값	249	1,021	1,241	309	327
	분산	63,504	5,324	5,447	4,912	5,017
음성군	추정값	2,214	1,306	1,454	2,208	2,209
	분산	744,769	4,893	5,066	4,861	5,032
보은군	추정값	915	573	624	915	914
	분산	763,876	1,542	1,575	1,539	1,572
옥천군	추정값	1,207	834	831	1,192	1,189
	분산	81,225	3,400	4,023	3,263	3,833
영동군	추정값	593	931	673	616	599
	분산	41,616	3,044	3,103	2,837	2,888
단양군	추정값	744	641	628	676	678
	분산	57,121	2,239	2,469	2,155	2,367
합계		29,742	29,742	29,742	29,742	29,742

표 3.4: 충북지역의 특성별 실업률

구분		남		여	
		시지역	군지역	시지역	군지역
연령대별	15-24	0.1244	0.0551	0.0515	0.1163
	25-34	0.0895	0.0817	0.0454	0.0177
	35-44	0.0668	0.0283	0.0545	0.0081
	43-54	0.0201	0.0325	0.0387	0.0173
	55세이상	0.0334	0.0094	0.0240	0.0089
교육 정도별	초졸	0.0427	0.0060	0.0161	0.0055
	중졸	0.0564	0.0171	0.0428	0.0219
	고졸	0.0684	0.0542	0.0510	0.0473
	대졸	0.0697	0.0723	0.0506	0.0000

$$\hat{y}_i = \sum_{j=1} x_{ij} r_j^a \quad (3.1)$$

단, i 는 시군구를 나타내고, j 는 성별-연령대별 분류를 나타낸다. x_{ij} 는 i 시군구의 j 성별-연령대의 경제활동 인구수를 나타내며 주민등록인구수에서 추정하거나 경찰조사에서 수집할 수 있으나 여기서는 후자를 이용하였고, r_j^a 는 j 성별-연령대의 실업률을 나타내는 것으로, j 성별-연령대의 경제활동 인구수에 대한 j 성별-연령대의 실업자수를 이용하여 계산할 수 있다. 즉,

$$r_j^a = \frac{y_j}{\sum_{i=1} x_{ij}}, \quad y_j: j \text{ 성별-연령대의 실업자수}$$

의 관계로부터 구할 수 있다. 이를 통해 계산된 1999년 4월의 충북지역의 특성별 실업률은 표3.4과 같다.

2. 성별-교육정도별 구성비를 이용한 시군구별 실업자 추정

교육정도별 구분은 초등학교졸, 중학교졸, 고등학교졸과 대학교졸(전문학교 포함)로 4개 범주로 나누어 실업률 추정의 정도를 높였다.

$$\hat{y}_i = \sum_{j=1} x_{ij} r_j^b \quad (3.2)$$

단, i 는 시군구를 나타내고, j 는 성별-교육정도별 분류를 나타낸다. x_{ij} 는 i 시군구의 j 성별-교육정도별 경제활동 인구수를 나타내고, r_j^b 는 j 성별-교육정도의 실업률을 나

타내는 것으로, j 교육정도의 경제활동 인구수에 대한 j 성별-교육정도의 실업자수를 이용하여 계산할 수 있다. 즉,

$$r_j^b = \frac{y_j}{\sum_{i=1} x_{ij}}, \quad y_j : j \text{ 성별-교육정도의 실업자수}$$

의 관계로부터 계산할 수 있으며, 그 결과는 표3.4과 같다.

식 (3.1)와 (3.2)으로 주어진 추정량의 분산은 각 범주별 실업률을 상수로 가정하고 x_{ij} 의 일차결합의 분산식을 통해서 계산하였다.

3.2.3. 복합 추정량

직접 추정량은 표본 수가 적기 때문에 분산이 클 뿐 만 아니라 표본의 크기에 민감하게 변동하고 합성 추정량은 각 범주들이 모든 시군구에서 동일하다는 전제 조건이 어긋날 경우에는 편향이 커지기 때문에 두 가지 추정량의 문제점을 완화하기 위해서 두 추정값의 가중 평균 형식인 복합 추정량을 생각하게 되었다.

$$\hat{Y}_i^c = w_i \hat{Y}_i^d + (1 - w_i) \hat{Y}_i^s \tag{3.3}$$

여기서 \hat{Y}_i^d 는 경찰 조사에서 직접 추정한 i 시군구의 실업자 수이고 \hat{Y}_i^s 는 성별-연령대별 분류 또는 성별-교육 정도별 분류에 의해서 합성 추정법으로 추정한 i 시군구의 실업자 수이다. w_i 는 $Var(\hat{Y}_i^d)$ 와 $mse(\hat{Y}_i^s)$ 에 의해서 계산되는 가중값으로 아래와 같이 표현할 수 있다.

$$w_{i(opt)} = \frac{mse(\hat{Y}_i^s)}{mse(\hat{Y}_i^s) + Var(\hat{Y}_i^d)}$$

여기서 $mse(\hat{Y}_i^s)$ 의 계산은 복잡하고 어려우므로 $Var(\hat{Y}_i^s)$ 로 대체하여 w_i 의 근사값을 다음과 같이 추정하였다.

$$\hat{w}_{i(opt)} = \frac{\widehat{Var}(\hat{Y}_i^s)}{\widehat{Var}(\hat{Y}_i^s) + \widehat{Var}(\hat{Y}_i^d)}$$

식 (3.3) 에서 주어진 추정량의 분산은 근사적으로 아래 식으로 계산할 수 있다.

$$\widehat{Var}(\hat{Y}_i^c) = \frac{\widehat{Var}(\hat{Y}_i^d) \cdot \widehat{Var}(\hat{Y}_i^s)}{\widehat{Var}(\hat{Y}_i^s) + \widehat{Var}(\hat{Y}_i^d)} \tag{3.4}$$

3.2.4. 추정된 실업자수 조정

성별-연령대별/성별-교육정도별의 각 범주의 실업률을 이용하여 계산된 시 지역의 실업자 합계는 경찰 조사 자료에서 직접 추정한 시 지역의 실업자 총수와 같다고 가정하였다. 이를 기준으로 각 시 구의 실업자 추정값을 다음과 같이 조정할 수 있다.

$$\hat{Y}_{i(A)} = \frac{\hat{Y}_i}{\sum_{i=1}^4 \hat{Y}_i} \hat{Y}^*$$

여기서 \hat{Y}_i 는 직접 추정법과 합성 추정법 및 복합 추정법으로 계산된 i 시·구의 실업자 수이고, \hat{Y}^* 는 직접 추정한 시 지역 실업자 총 수이다.

군 지역의 실업자 총수는 통계청에서 추정된 충청북도의 실업자수에서 시·구 지역의 조정된 실업자 수를 감하여 계산하고 이 결과를 이용하여 각 군별로 실업자 추정값을 다음과 같이 조정한다. 즉 군지역의 조정된 실업자 추정값은

$$\hat{Y}_{i(A)} = \frac{\hat{Y}_i}{\sum_{i=1}^9 \hat{Y}_i} \hat{Y}^{**}$$

로 구한다. 여기서 \hat{Y}_i 는 직접 추정법과 합성 추정법 및 복합 추정법에서 추정된 군 지역의 실업자 수이고, $\hat{Y}^{**} = (\text{충북실업자수} \cdot \text{시·군의 실업자합계})$ 에 대한 추정값이다.

3.3. 추정결과

표3.3에 주어진 추정값과 분산에서 특징은 직접 추정량의 분산이 가장 크고 합성 추정량의 분산은 편향을 생략하였기 때문에 직접 추정량보다는 작을 것으로 생각했지만 13배에서 200배 정도까지 감소된 점은 표본의 크기가 작을 경우에는 합성 추정법이 직접 추정법에 비해서 월등하게 효과적임을 설명하고 있다.

합성 추정법과 복합 추정법의 비교에서 복합 추정량이 안정적인 특성을 보이고 있을 뿐만 아니라 분산도 약간 감소함을 볼 수 있다. 3가지 추정법 중에서는 복합 추정법이 월등함을 수치적인 사례로 보여주고 있다.

4. 결론

지방 자치제의 정착으로 각 시군구의 경제 활동 관련 통계의 필요성이 제기되었을 뿐만 아니라 실업 대책 및 공공근로사업의 예산 배분 등으로 적절한 예산 배정의 기준 설정을 위해서 시군구의 경제활동인구, 취업자 및 실업자 수에 대한 국회와 행자부의 요청이 있었으나 합당한 방법이 없었다. 그래서 몇몇 시도에서 자체적인 통계 조사 계획으로 시군구 실업자 수를 추진하고자 시도하였으나 신뢰성 있는 결과를 얻을 수 없었다.

시군구의 실업자 및 경제활동관련 통계는 필수적으로 작성되어야 하므로 많은 예산과 인원을 투입하지 않고도 시군구의 통계 생산이 가능한 소지역 추정법에 대한 연구는 시급하게 추진되어야 할 것이다.

소지역 추정법은 미국이나 캐나다에서는 오래 전부터 활용되어 왔으며 영국이나 프랑스 등 유럽에서는 국가적인 차원에서 연구가 진행중이다.

본 연구에서는 지금까지 연구되어 활용되고 있는 내용을 중심으로 소지역 추정법을 요약했고 충북 시군구의 실업자 추정에 합성과 복합 추정법을 적용시켜 본 결과에서 안정적이고 손쉽게 시군구의 실업통계를 생산할 수 있음을 보여 주고 있다.

본 연구는 이제 시작일 뿐이다. 좀더 신뢰성 있고 광범위한 정보(실업 보험 신청 dB, 구직 등록 dB)를 이용할 수 있는 방안과 경향 조사가 매월 패널 조사로 이루어지고 있으므로 이들의 시계열 특성을 이용하여 각 지역별로 모형을 추정하여 베이지안 추정법의 적용 방

안에 대한 연구를 추진해야만 2000년 센서스 후 경활 조사 표본을 개편할 때 소지역 추정법을 적용할 수 있을 것이다.

앞으로 정보화 시대에서는 정부 부처간의 유관 정보는 공유할 수 있는 체제도 함께 추진해야만 신뢰성있는 정부 통계를 생산하는데 도움을 줄뿐 만 아니라 점점 어려워지는 조사 환경에서 신속 정확한 통계의 생산에도 도움이 될 것이다.

참고문헌

- [1] 이계오, 류재복, 이규형 (1999). 소지역 추정법을 이용한 충북 시군구 실업자 추정, <“소지역 통계” 워크샵>, 1999.6.29, 37-56.
- [2] 통계청 (1999). <경제활동 인구 월보 1999.4-5>
(<http://WWW.stat.go.kr>)
- [3] 통계청 (1996.8). <표본개편 연구회 연구보고>
- [4] Ghosh, M. and Rao, J.N.K. (1994). Small Area Estimation : An Appraisal, *Statistical Science*, Vol. 9, 55-93.
- [5] Gonzalez, M.E. and Hoza, C. (1978). Small Area Estimation with Application to Unemployment and Housing Estimates, *Journal of the American Statistical Association*, Vol. 73, 7-15.
- [6] Purcell, N.J. and Kish, L. (1979). Estimation for Small Domain, *Biometrics*, Vol. 35, 365-384.
- [7] Wolter, K.M. (1985). *Introduction to Variance Estimation*, Springer-Verlag, New York.

[2000년 4월 접수, 2000년 6월 채택]

On Application of Small Area Estimation to the Unemployment Statistics of Si-Gun-Gu

Kay-O Lee ¹⁾

ABSTRACT

After commencing the local self-governing system, it has been requested to make statistics about Si-Gun-Gu level, and the necessity of unemployment statistics in Si-Gun-Gu area has also been proposed as the unemployment became a political issue and social problem due to an IMF situation. Throughout this paper, therefore, we suggest two small area estimations that can produce the unemployment statistics of Si-Gun-Gu based on the Current Population Survey. In addition, we estimate the unemployment number of Si-Gun-Gu in ChungBuk area from the real data of the Current Population Survey. Finally, It was also proved that the applicability of small area estimation was valid in comparison with efficiency of small area estimations.

Keywords: Small Area Estimation; Synthetic Estimator; Composite Estimator.

1) Professor, Department of Computer Science and Statistics, Korea Air Force Academy.
E-mail: kayolee@hanmail.com