

연구발표자료

공군사관학교 전산통계학과

이 계 오 교수

소지역추정법을 이용한
시군구 실업자 추계연구

2001. 10.

통계기획국 조사관리과

목 차

1. 서 론	2
2. 소지역 추정법	4
3. 실증조사 결과분석	13
4. 시·군·구 실업자 추정과제	24
5. 결 언	25

1. 서 론

- 1995년 지방자치제도의 실시로 시군구 단위 통계생산 필요
- IMF 사태를 겪으면서 고용관련 통계에 대한 관심 증대
- 정보화 사회에서 통계정보의 공유와 활용성이 강조되면서 신속 정확한 통계의 생산이 요구됨

※ 시군구 실업자 추계에 대한 연구의 필요성 대두

◎ 현재의 경제활동인구조사 실태

- 광역시와 도단위의 경제활동관련 통계작성을 목표로 표본을 설계함
- 1999년 연간 경제활동인구 통계의 상대표준오차

<표 1.1> 경제활동인구조사의 상대표준오차

시·도	경제활동인구	취업자	실업자	비경제활동인구
전국	0.12	0.13	0.79	0.19
서울	0.30	0.29	1.86	0.46
부산	0.39	0.43	2.11	0.54
대구	0.48	0.52	2.23	0.65
인천	0.37	0.40	2.18	0.56
광주	0.44	0.46	2.18	0.52
대전	0.40	0.43	3.08	0.50
울산	0.73	0.76	3.82	1.03
경기	0.32	0.35	2.13	0.52
강원	0.54	0.58	3.76	0.75
충북	0.58	0.62	3.12	0.84
충남	0.53	0.57	3.54	0.92
전북	0.54	0.58	3.34	0.75
전남	0.52	0.58	3.08	0.97
경북	0.47	0.46	3.60	0.85
경남	0.59	0.61	3.88	0.90
제주	0.65	0.65	5.79	1.15

<표 1.1>에 있는 지역별 주요 항목의 상대표준오차는 1999년 연간 기준의 상대표준오차이므로 월단위 추정값 기준의 상대표준오차는 훨씬 커질 것이다. 실업자 추정값의 연가 기준의 상대표준오차는 제주도를 제외하면 4% 이내이므로 문표

허용오차를 만족하고 전국의 표본규모는 예산과 조사인력에 근거하여, 약 30,000 가구로 한정된 후에 지역별로 조사인력과 상대표준오차를 고려하여 표본조사구를 배분하였다.

- 시도별 표본 조사구 배분 현황

<표 1.2 > 지역별 표본조사구 배분

지역	표본조사구	지역	표본조사구
서울	160	강원	71
부산	92	충북	64
대구	72	충남	70
인천	93	전북	63
광주	87	전남	67
대전	65	경북	71
울산	30	경남	62
경기	118	제주	34

- 광주광역시와 충북의 시군구 조사구 분포

구이름	동 구	서 구	남 구	북 구	광산구	합계
조사구 수	10	15	16	34	12	87

시군명	청주시	충주시	제천시	청원군	진천군	괴산군
조사구 수	22	11	4	5	1	5
시군명	음성군	단양군	옥천군	보은군	영동군	합계
조사구 수	6	2	3	2	3	64

※ 시군구 실업자 추정을 위한 소지역 추정법에 대한 연구 필요함

2. 소지역 추정법

소지역 통계 추정을 위한 추정법은 크게 설계 기반 추정법(design-based estimation), 간접추정법(indirect estimation), 모형 기반 추정법(model-based estimation)으로 구분할 수 있다. 소지역 통계 작성 시 설계 기반 추정량이 목표 요구정도를 만족한다면 우선적으로 설계 기반 추정량을 이용하며 그렇지 못할 경우에는 추정량의 신뢰도를 확보할 수 있는 다른 추정법을 찾아야한다.

(1) 설계 기반 추정법(Design-Based Estimation)

일반적으로 설계 기반 추정법에서는 직접추정량(direct estimator)과 수정된 직접추정량(modified direct estimator)으로 구분할 수 있다. 관심변수와 밀접한 관련이 있는 보조정보가 있는 경우에 이를 이용하는 사후층화추정량(post stratified estimator), 비추정량(ratio estimator), 회귀추정량(regression estimator) 등은 직접추정량의 일종이다. 직접추정량은 편향이 없는 추정량이지만 해당 소지역에 배정된 표본의 크기가 작은 경우에는 추정량의 분산이 커져서 신뢰성이 떨어지게 된다. 한편 수정된 직접추정량(modified direct estimator)은 해당 소지역 이외의 다른 지역의 조사결과를 추정과정에 추가적으로 이용하며 추정량의 불편성은 근사적으로 유지된다.

직접추정량(direct estimator)은 보통 해당 소지역에서 조사된 자료만을 이용하여 추정되며, 간혹 센서스나 행정자료로부터 획득된 보조정보를 조사자료에 추가하여 추정하기도 한다. 가장 간단한 총계추정에 대한 직접추정량으로써 다음과 같은 단순추정량(expansion estimator)을 들 수 있다.

$$\hat{Y}_{e,a} = \sum_{i \in s_a} \omega_i y_i \quad (2.1)$$

여기에서 s_a 는 소지역 a 의 표본들의 집합, ω_i 는 조사단위 i 에 대한 가중치를 나타낸다. 위의 직접추정량은 불편추정량이나 소지역 a 의 표본크기가 작을 경우에는 분산이 커지기 때문에 신뢰성에 문제가 있을 수 있다.

소지역 a 의 모집단의 크기 N_a 를 알고있을 경우에는 다음과 같은 사후층화추정량이 이용될 수 있다.

$$\hat{Y}_{pst,a} = N_a \frac{\sum_{i \in s_a} \omega_i y_i}{\sum \omega_i}$$

$$\begin{aligned}
&= N_a \frac{\hat{Y}_{e,a}}{\hat{N}_{e,a}} \\
&= N_a \bar{y}_{e,a} \quad (2.2)
\end{aligned}$$

위의 사후층화추정량은 먼저 언급된 단순추정량보다는 안정적이나 보다 복잡한 조사에서는 비추정편향(ratio estimation bias)이 발생할 가능성이 있다.

표본이 층화추출되고 층 h 에서 소지역 a 의 모집단의 크기 $N_{h,a}$ 가 알려져 있을 경우에는 다음과 같은 유형의 사후층화추정량이 소지역 추정에 이용될 수 있다.

$$\begin{aligned}
\hat{Y}_{st, pst, a} &= \sum_h \left(N_{h,a} \frac{\sum_{i \in S_{h,a}} \omega_i y_i}{\sum_{i \in S_{h,a}} \omega_i} \right) \\
&= \sum_h N_{h,a} \frac{\hat{Y}_{h,e,a}}{\hat{N}_{h,e,a}} \\
&= \sum_h N_{h,a} \bar{y}_{h,a} \quad (2.3)
\end{aligned}$$

여기에서 층 h 는 표본설계 시 반영된 층이라기 보다는 사후층화에 의해 형성된 층을 말한다.

비추정법(ratio estimation)은 사후층화추정법과 유사하나 모집단 총계 N_a 와 $N_{h,a}$ 대신에 보조정보에 의해 획득된 소지역 총계 X_a 와 $X_{h,a}$ 를 이용하며, 이 값들을 알고 있을 경우 비추정량은 다음과 같이 정의된다.

$$\hat{Y}_{r,a} = X_a \hat{R}_a, \quad \hat{Y}_{st,r,a} = \sum_h X_{h,a} \hat{R}_{h,a}, \quad (2.4)$$

여기에서 $\hat{R}_a = \hat{Y}_{e,a} / \hat{X}_{e,a}$ 는 Y_a / X_a 의 추정값, $\hat{R}_{h,a} = \hat{Y}_{h,e,a} / \hat{X}_{h,e,a}$ 를 나타낸다.

회귀추정법(regression estimation)이 소지역 총계 추정에 이용되기도 한다. 이 방법은 관심변수 y 와 공변량 x 사이의 관계에서 회귀모수를 추정하여 소지역 총계 추정에 이용하는 방법으로써 추정량은 다음과 같은 형태로 주어진다.

$$\hat{Y}_{reg,a} = \hat{Y}_a + \hat{\beta}_a (X_a - \hat{X}_a), \quad (2.5)$$

여기에서 \hat{Y}_a 는 직접추정 또는 사후층화추정법에 의해 추정된 소지역 a 에 대한

총계 추정값이며 \hat{X}_a 는 보조정보를 통해 \hat{Y}_a 과 유사한 방법으로 추정된다. 추정 모수 β_a 은 관심변수 y 와 공변량 x 의 관계로부터 추정되며 $\hat{\beta}_a = \sum_{i \in S_a} \nu_i^{-1} \omega_i y_i x_i^T (\sum_{i \in S_a} \nu_i^{-1} \omega_i x_i x_i^T)^{-1}$ 와 같이 주어진다. 여기에서 ν_i 는 회귀가중치로써 주어지는 값이며, x 가 상수이고 $\nu_i = x_i$ 일 경우에는 $\hat{\beta}_a = \hat{R}_a$ 인 관계가 성립한다. 회귀추정량의 불편성은 \hat{Y}_a 와 \hat{X}_a 의 불편성에 의존한다.

한편, 회귀추정량을 변형한 일종의 수정된 직접추정량(modified direct estimator)이 소지역 특성치 추정에 이용되기도 한다. 수정된 직접추정량은 해당 지역 외의 조사자료를 특성치 추정에 이용하며, 추정량의 불편성은 회귀추정량과 마찬가지로 근사적으로 만족된다. 예를 들면 식(7.5)에서 추정 회귀모수 $\hat{\beta}_a$ 대신에 회귀모수에 대한 합성추정량의 일종인 $\hat{\beta} = \sum_{i \in S} \nu_i^{-1} \omega_i y_i x_i^T (\sum_{i \in S} \nu_i^{-1} \omega_i x_i x_i^T)^{-1}$ 이 이용되었다면 이러한 추정량을 수정된 직접추정량(modified direct estimator)이라 부른다. 일반적으로 소지역 추정 시 $\hat{\beta}$ 이 $\hat{\beta}_a$ 보다 안정적인 것으로 알려져 있으며, $\hat{\beta}$ 과 $\hat{\beta}_a$ 의 가중평균 $\lambda_a \hat{\beta}_a + (1 - \lambda_a) \hat{\beta}$ 이 추정 회귀모수로 이용되기도 한다. 여기에서 λ_a 는 적절히 선택되는 값이다. x 가 상수이고 $\nu_i = x_i$ 인 경우에는 $\hat{\beta}$ 대신 $\hat{R} = \hat{Y}_e / \hat{X}_e$ 이 이용될 수도 있다.

(2) 간접 추정법(Indirect Estimation)

간접 추정법에서는 합성추정량(synthetic estimator), 복합추정량(composite estimator), 표본수 의존 복합추정량(sample size dependent estimator) 등의 유형이 다루어 질 것이며, 해당 지역의 조사자료뿐만 아니라 해당 지역을 포함하고 있는 더 큰 지역의 조사자료를 소지역 추정과정에 이용하여 소지역 추정의 신뢰성을 확보하는 방법이다.

합성추정법(synthetic estimation)은 소지역 추정 시 소지역을 포함하는 대영역의 정보를 함께 이용하는 방법으로써 소지역과 대영역의 특성 구조가 유사하다는 가정 하에서 이용된다. 합성추정량의 분산은 직접추정량의 분산에 비해 작으나 전제한 가정이 성립하지 않을 경우에는 심각한 편향이 발생할 수 있다.

소지역의 특성치 평균이 전체 지역의 특성치 평균과 같다는 가정 하에서 만들어진 가장 간단한 형태의 합성추정량은 다음과 같다.

$$\hat{Y}_{syn, m, a} = N_a \frac{\sum_{i \in S} \omega_i y_i}{\sum_{i \in S} \omega_i} = N_a \bar{y} \quad (2.6)$$

층화 또는 사후층화에 근거한 합성추정량은 보통 다음과 같은 형태로 주어진다.

$$\hat{Y}_{syn, st, m, a} = \sum_h N_{h, a} \frac{\sum_{i \in S_h} \omega_i y_i}{\sum_{i \in S_h} \omega_i} = \sum_h N_{h, a} \bar{y}_h \quad (2.7)$$

직접추정법에서와 마찬가지로 합성추정법에서도 비합성추정법(ration synthetic estimation)이 고려될 수 있다. 비합성추정법은 모집단의 크기 N_a 또는 $N_{h, a}$ 외의 소지역 추정을 위한 보조정보로써 공변량 x 를 이용하며 추정량(ratio synthetic estimator)은 다음과 같은 형태로 정의된다.

$$\hat{Y}_{syn, r, a} = X_a \frac{\hat{Y}_e}{\hat{X}_e}, \quad \hat{Y}_{syn, st, r, a} = \sum_h X_{h, a} \frac{\hat{Y}_{h, e}}{\hat{X}_{h, e}}, \quad (2.8)$$

여기에서 $\hat{Y}_e = \sum_{i \in S} \omega_i y_i$ 는 y 에 대한 모집단 총계 추정량, $\hat{Y}_{h, e} = \sum_{i \in S_h} \omega_i y_i$ 는 층의 총계 추정치를 나타낸다. 비합성추정량들은 Gonzalez(1973), Gonzalez and Waksberg(1973), Ghangurde and Singh(1977, 1978)에 자세히 소개되어있다. 한편, Singh and Tessier(1976)는 (2.8)식의 $\hat{Y}_{syn, r, a}$ 에서 \hat{X}_e 대신에 X 를 이용한 비합성 추정량의 대체식 $\hat{Y}_{syn, r, a} = X_a \hat{Y}_e / X$ 를 제안하였다. 여기에서 $\hat{Y}_{syn, r, a}$ 와 $\hat{Y}_{syn, st, r, a}$ 는 모두 같은 양의 편향을 가지며, $\hat{Y}_{syn, r, a}$ 의 편향은 표본의 크기가 클 경우에는 무시될 수 있다. 두 추정량 중 하나의 추정량을 선택하는 문제에서는 \hat{Y}_e 와 \hat{X}_e 의 상관계수 ρ 를 참조하도록 하였다. 일반적으로 표본의 크기가 클 경우, 두 추정량의 분산은 상관계수의 값이 $\rho \geq 0.5 c_x / c_y$ 일때 $V(\hat{Y}_{syn, r, a}) \leq V(\hat{Y}_{syn, st, r, a})$ 인 관계가 성립한다. 여기에서 c_x 와 c_y 는 각각 \hat{X}_e 와 \hat{Y}_e 의 변동계수(coefficient of variation)를 나타낸다. 상관계수 ρ 의 값이 크거나 모집단의 분포가 한쪽으로 치우쳐져 있을 경우에는 $\hat{Y}_{syn, r, a}$ 가 선호되며, 변동계수 c_x 의 값이 크거나 상관계수 ρ 의 값이 적당할 경우에는 보통 $\hat{Y}_{syn, st, r, a}$ 를 선택한다.

소지역의 보조정보로써 이용된 공변량 x 외에 추가적인 보조변수 z 를 도입하여 소지역 특성치를 추정하는 다음과 같은 이변량 비합성추정량이 소지역 추정에 이용

될 수 있다.

$$\hat{Y}_{syn, r, a}^{(2)} = \gamma_a X_a \frac{\hat{Y}_e}{\hat{X}_e} + (1 - \gamma_a) Z_a \frac{\hat{Y}_e}{\hat{Z}_e}, \quad (2.9)$$

여기에서 γ_a 는 적절히 선택되는 값이다. 보다 일반적인 다변량 비합성추정량은 Olkin(1958)에서 참조할 수 있다.

회귀합성추정법은 비합성추정법과 유사하며 추정량은 다음과 같이 주어진다.

$$\hat{Y}_{syn, reg, a} = \hat{\beta} X_a, \quad \hat{\beta} = \sum_{i \in S_a} \nu_i^{-1} \omega_i y_i x_i^T \left(\sum_{i \in S_a} \nu_i^{-1} \omega_i x_i x_i^T \right)^{-1} \quad (2.10)$$

회귀합성추정법은 표본설계의 층 내에서 또는 사후층화에 의해 형성된 층 내에서도 응용이 가능하며, Royall(1979)은 이러한 내용을 반영한 다음과 같은 회귀합성추정량을 제안하였다.

$$\hat{Y}_{syn, Roy, a} = \sum_{i \in S_a} y_i + \hat{\beta} (X_a - \sum_{i \in S_a} x_i), \quad (2.11)$$

복합추정량(composite estimator)은 직접추정량의 불안정성과 합성추정량의 잠재적 편향 가능성을 보완하기 위해 두 추정량의 가중평균을 취하며 일반적인 형태는 다음과 같이 주어진다.

$$\hat{Y}_{com, a} = \lambda_a \hat{Y}_{dir, a} + (1 - \lambda_a) \hat{Y}_{syn, a}, \quad (2.12)$$

여기에서 가중치 λ_a 는 적절히 선택되는 값이다.

가중치 λ_a 는 결정하는 방법은 크게 세가지 정도로 구분될 수 있다. 첫 번째 방법은 가장 간단한 방법으로써 가중치 λ_a 를 고정계수로 두는 방법인데 추정량의 신뢰성에 문제가 있어 많이 사용되지는 않는다. 두 번째 방법은 추정하고자하는 소지역의 표본크기를 반영하는 방법이다. 이 경우 가중치 λ_a 는 $\hat{N}_{e, a}/N_a$ 의 함수로 표현된다. Drew et al.(1982)은 표본크기에 의존하는 복합추정량으로써 다음과 같은 추정량을 제안하였다.

$$\hat{Y}_{ssd, r, a} = \lambda_a \hat{Y}_{r, a} + (1 - \lambda_a) \hat{Y}_{syn, r, a}, \quad (2.13)$$

여기에서 $\lambda_a = \begin{cases} 1 & , \text{ if } \hat{N}_{e,a} \geq \delta N_a \\ \hat{N}_{e,a} / \delta N_a & , \text{ otherwise} \end{cases}$ 이며, δ 는 합성추정량 부분의 편향을

보정하기 위해 주관적으로 결정되는 값이다. 캐나다 노동력조사에서는 $\delta = 2/3$ 를 이용한다. 식 (2.13)의 복합추정량은 직접추정량의 신뢰도가 완전히 확보된 지역에 대해서는 합성추정량의 가중치가 0이 되기 때문에 이러한 지역의 경우 직접추정값이 곧 복합추정값으로 선택된다고 볼 수 있다. 그렇지 않은 기타 지역에 대해서는 직접추정값과 합성추정값의 가중평균값으로 복합추정값이 계산된다. 캐나다 노동력 조사에서 이러한 기타 지역들에 대한 합성추정량의 평균 가중치는 약 10% 정도이며 많아야 20%를 초과하지는 않는다. 이때 δ 의 값은 $[2/3, 3/2]$ 의 범위에 있는 것으로 알려져 있다. 이외의 표본크기 의존 복합추정량으로써 Sandal(1984)의 추정량

$\hat{Y}_{ssd, reg, a} = \lambda_a \hat{Y}_{sreg, a} + (1 - \lambda_a) \hat{Y}_{syn, reg, a}$ 을 들 수 있다. 사용된 가중치는

$\lambda_a = \hat{N}_{e,a} / N_a$ 이다. Rao(1986)는 위와 동일한 추정량에 대해 가중치를 약간 달리

적용할 것을 제안하였다. Rao의 가중치는 $\hat{N}_{e,a} \geq N_a$ 인 지역에 대해서는 $\lambda_a = 1$,

기타 지역에 대해서는 Sandal의 가중치와 동일하다. Sandal and Hidiroglou(1989)는

Rao의 가중치에서 $\hat{N}_{e,a} < N_a$ 일때 $\lambda_a = (\hat{N}_{e,a} / N_a)^{h-1}$ 를 사용할 것을 제안하였다.

여기에서 h 는 합성추정량 편향을 감안하여 적절히 선택되는 값이다. 가중치를 결정하는

세 번째 방법은 직접추정량과 합성추정량의 평균제곱오차와 두 추정량의 공분산을

자료로부터 추정하여 최적가중치를 산정하는 방법이다. 복합추정량의 평균

제곱오차는 다음 식과 같이 나타낼 수 있다.

$$MSE(\hat{Y}_{com, a}) = \lambda_a^2 MSE(\hat{Y}_{dir, a}) + (1 - \lambda_a)^2 MSE(\hat{Y}_{syn, a}) + 2\lambda_a(1 - \lambda_a)E(\hat{Y}_{dir, a} - Y_a)(\hat{Y}_{syn, a} - Y_a) \quad (2.14)$$

$\hat{Y}_{com, a}$ 의 MSE를 최소화하는 가중치 λ_a 다음 식과 같이 주어질 수 있다.

$$\lambda_a = \frac{\widehat{MSE}(\hat{Y}_{syn, a}) - E(\hat{Y}_{syn, a} - Y_a)(\hat{Y}_{dir, a} - Y_a)}{\widehat{MSE}(\hat{Y}_{syn, a}) + \widehat{MSE}(\hat{Y}_{dir, a}) - 2E(\hat{Y}_{syn, a} - Y_a)(\hat{Y}_{dir, a} - Y_a)} \quad (2.15)$$

식 (2.15)에서 $\hat{Y}_{dir, a}$ 와 $\hat{Y}_{syn, a}$ 의 공분산의 항이 $\widehat{MSE}(\hat{Y}_{syn, a})$ 와 $\widehat{MSE}(\hat{Y}_{dir, a})$

에 비해 매우 작다고 가정할 수 있다면 다음과 같은 근사적인 가중치를 이용할 수

도 있다.

$$\lambda_a^* = \frac{\widehat{MSE}(\hat{Y}_{syn,a})}{\widehat{MSE}(\hat{Y}_{syn,a}) + \widehat{MSE}(\hat{Y}_{dir,a})} \quad (2.16)$$

(3) 모형 기반 추정법(Model-Based Estimation)

소지역 추정에 자주 이용되는 모형 기반 추정법(model-based estimation)으로는 EBLUP (empirical best linear unbiased prediction), EB(empirical Bayes), HB(hierarchical Bayes) 접근법 등이 있다. 최근에는 소지역 통계 작성을 위해 횡단면 조사자료(cross-sectional data)와 시계열자료(time series data)를 함께 추정과정에 이용하는 방법에 관한 연구가 활발히 진행되고 있다. 횡단면 조사자료를 이용한 모형기반 추정량과 횡단면 조사자료와 시계열자료를 함께 이용하는 모형기반 추정량에 대한 연구들이 최근에 진행되고 있으나 이 절에서는 시계열 자료와 결합하는 모형은 설명하지 않겠다.

y_i 를 i 번째 소지역의 관심모수 θ_i 에 대한 직접추정량, x_i 를 모수 θ_i 의 추정에 필요한 설명변수이고, 모형 $y_i = \theta_i + e_i$, $E(e_i) = 0$ 를 가정할때, 소지역 i 에 대한 다음과 같은 선형회귀모형(linear regression model)을 고려할 수 있다.

$$\theta_i = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, I, \quad (2.17)$$

여기에서 β_0 와 β_1 은 회귀모수를 나타낸다. 이때 θ_i 에 대한 회귀합성추정량(regression synthetic estimator)은 다음과 같이 주어질 수 있다.

$$\hat{\theta}_{i(reg)} = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, 2, \dots, I, \quad (2.18)$$

여기에서 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 은 결합모형 $y_i = \beta_0 + \beta_1 x_i + e_i$ ($i = 1, 2, \dots, I$)로부터 계산되는 최소제곱추정량을 나타낸다. 조사 추정량 y_i 들의 공분산을 추정할 수 있을 경우에는 일반화 가중최소제곱추정량을 이용할 수도 있다. 위의 회귀합성추정량은 조사 추정량 y_i 들에 대한 가중치가 반영되지 않기 때문에 큰 편향이 발생할 수 있다. 반면, EB(Empirical Bayes) 추정량이나 EBLUP(Empirical Bayes Linear Unbiased Predictor)는 적당한 가중치가 부여되어 편향 발생이 다소 억제되는 결과를 얻을 수 있다.

이러한 편향에 대한 문제점을 해결하기 위해 Fay and Herriot(1979)는 모형 (2.17)을 다음과 같이 해당 소지역에 대한 랜덤효과 v_i 를 갖는 모형으로 보완하였

($T \times T$ 행렬)로 가정한다. 모평균 θ_{it} 에 관한 모형은 다양한 유형으로 설정될 수 있으며 다음과 같은 모형들이 고려될 수 있다.

$$(I) \theta_{it} = \beta_0 + \beta_1 x_{it} + v_i + \varepsilon_{it}, \quad i=1, 2, \dots, I, \quad t=1, 2, \dots, T,$$

여기에서 v_i 는 소지역 고정효과, $\varepsilon_{it} \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$.

$$(II) \theta_{it} = \beta_0 + \beta_1 x_{it} + v_i + \varepsilon_{it}, \quad i=1, 2, \dots, I, \quad t=1, 2, \dots, T,$$

여기에서 $v_i \stackrel{\text{ind}}{\sim} N(0, \sigma_v^2)$, $\varepsilon_{it} \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$, $\{v_i\}$ 와 $\{\varepsilon_{it}\}$ 는 서로 독립이며, 모

형(I)과는 달리 v_i 들이 랜덤효과로 가정되었다.

$$(III) \theta_{it} = \beta_0 + \beta_1 x_{it} + v_i + u_t + \varepsilon_{it}, \quad i=1, 2, \dots, I, \quad t=1, 2, \dots, T,$$

여기에서 $v_i \stackrel{\text{ind}}{\sim} N(0, \sigma_v^2)$, $u_t \stackrel{\text{ind}}{\sim} N(0, \sigma_u^2)$, $\varepsilon_{it} \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$, $\{v_i\}$, $\{u_t\}$ 와

$\{\varepsilon_{it}\}$ 는 서로 독립이다. v_i 는 소지역에 대한 랜덤효과, u_t 는 조사시기에 대한 랜덤효과이다.

3. 실증 조사 결과 분석

3.1 실증조사의 조사구와 자료

현행 경찰자료를 이용하여 추계한 소지역 추정값의 타당성을 수치적으로 입증하기 위해서 광주 광역시에서 14개조사구, 충북에서 74개조사구를 추가하여 시군구의 실업통계를 작성 비교분석함

광주광역시의 북구와 충북 청주시는 기존 조사구가 충분히 큰 것으로 생각되어 조사구를 추가하지 않고, 기타 시군구는 최소한 10개조사구 이상을 조사토록 함.

<표 3.1> 시군구별 조사구 분포와 주민등록인구

시도	시군구	조사구 수			주민등록인구
		기존	추가	합계	
광주 광역시	동구	10	4	14	123,647
	서구	15	3	18	287,867
	남구	16	2	18	231,501
	북구	34	0	34	475,992
	광산구	12	5	17	252,902
	소계	87	14	101	1,371,909
충청 북도	청주시	22	0	22	580,861
	청원군	5	9	14	123,984
	진천군	1	9	10	60,121
	괴산군	5	6	11	76,666
	충주시	11	5	16	217,305
	제천시	4	10	14	147,950
	음성군	6	5	11	87,956
	단양군	2	8	10	40,000
	옥천군	3	7	10	60,798
	보은군	2	8	10	43,245
	영동군	3	7	10	58,627
	소계	64	74	138	1,497,513
합계		151	88	239	

광주광역시의 표본조사 자료를 요약하였으며 직접 추정값을 계산하기 위한 승수는 기존 조사구는 남자/여자 : 218.265/203.010 이고 추가 조사구는 남자/여자 : 189.058/175.299이다.

<표3.2> 광주광역시 5월 구별 자료요약

구분	경제활동인구			취업자			실업자			비경제활동인구			
	기존	추가	합계	기존	추가	합계	기존	추가	합계	기존	추가	합계	
동구	남자	141	63	204	130	59	189	11	4	15	99	44	143
	여자	144	40	184	142	38	180	2	2	4	145	73	218
	계	285	103	388	272	97	369	13	6	19	244	117	361
서구	남자	262	45	307	251	41	292	11	4	15	158	26	184
	여자	200	38	238	188	37	225	12	1	13	252	50	302
	계	462	83	545	439	78	517	23	5	28	410	76	486
남구	남자	255	38	293	241	33	274	14	5	19	130	16	146
	여자	234	26	260	223	23	246	11	3	14	233	36	269
	계	489	64	553	464	56	520	25	8	33	363	52	415
북구	남자	563	-	-	543	-	-	20	-	-	281	-	-
	여자	475	-	-	454	-	-	21	-	-	559	-	-
	계	1038	-	-	997	-	-	41	-	-	840	-	-
광산구	남자	241	84	325	233	82	315	8	2	10	64	53	117
	여자	167	66	233	162	64	226	5	2	7	151	85	236
	계	408	150	558	395	146	541	2	4	17	215	138	353
합계	남자	1462	230	1692	1398	215	1613	64	15	79	732	139	871
	여자	1220	170	1390	1169	162	1331	51	8	59	1340	244	1584
	계	2682	400	3082	2567	377	2944	115	23	138	2072	383	2455

충북의 5월 조사자료를 아래와 같이 요약하였고 직접 추정값의 승수는 시
 지역 남자/여자 : 389.188 / 362.123 (289.303 / 272.007)이고 읍면지역은 남
 자/여자 : 263.218 / 252.119 (89.485 / 85.798)이다.

<표3.3> 충청북도 5월 시군별 자료요약

구분	경제활동인구			취업자			실업자			비경제활동인구			
	기존	추가	합계	기존	추가	합계	기존	추가	합계	기존	추가	합계	
청주시	남	375	0	375	356	0	356	19	0	19	192	0	192
	여	293	0	293	286	0	286	7	0	7	339	0	339
	계	668	0	668	642	0	642	26	0	26	531	0	531
제천시	남	63	145	208	62	142	204	1	3	4	21	72	93
	여	46	96	142	45	93	138	1	3	4	55	143	198
	계	109	241	350	107	235	342	2	6	8	76	215	291
충주시	남	209	102	311	203	100	303	6	2	8	67	68	135
	여	185	58	243	182	58	240	3	0	3	131	86	217
	계	394	160	554	385	158	543	9	2	11	198	154	352
보은군	남	40	150	190	39	145	184	1	5	6	4	50	54
	여	29	123	152	29	120	149	0	3	3	28	100	128
	계	69	273	342	68	265	333	1	8	9	32	150	182
옥천군	남	79	139	218	76	134	210	3	5	8	15	51	66
	여	69	108	177	69	106	175	0	2	2	34	102	136
	계	148	247	395	145	240	385	3	7	10	49	153	202
영동군	남	46	138	184	46	136	182	0	2	2	27	36	63
	여	57	115	172	56	115	171	1	0	1	33	85	118
	계	103	253	356	102	251	353	1	2	3	60	121	181
괴산군	남	91	100	191	89	97	186	2	3	5	21	22	43
	여	84	87	171	83	87	170	1	0	1	54	63	117
	계	175	187	362	172	184	356	3	3	6	75	85	160
음성군	남	128	99	227	125	98	223	3	1	4	43	38	81
	여	92	75	167	90	73	163	2	2	4	72	68	140
	계	220	174	394	215	171	386	5	3	8	115	106	221
청원군	남	91	179	270	90	172	262	1	7	8	35	64	99
	여	79	138	217	77	137	214	2	1	3	84	123	207
	계	170	317	487	167	309	476	3	8	11	119	187	306
진천군	남	25	187	212	23	184	207	2	3	5	9	61	70
	여	20	137	157	19	136	155	1	1	2	15	137	152
	계	45	324	369	42	320	362	3	4	7	24	198	222
단양군	남	53	154	207	53	149	202	0	5	5	6	53	59
	여	41	106	147	41	106	147	0	0	0	23	108	131
	계	94	260	354	94	255	349	0	5	5	29	161	190
합계	남	1200	1393	2593	1162	1357	2519	38	36	74	440	515	955
	여	995	1043	2038	977	1031	2008	18	12	30	868	1015	1883
	계	2195	2436	4631	2139	2388	4527	56	48	104	1308	1530	2838

3.3 실업자 추정

(1) 직접 추정량

$$\begin{aligned}
 \hat{Y}_{i\cdot} &= \sum_{s=1}^2 {}_s\hat{Y}_{i\cdot} \quad , \quad i=1,2,\dots,I ; s=1,2 ; h=1,2,\dots,n_i \\
 &= \sum_{s=1}^2 \sum_{h=1}^{n_i} {}_s\hat{Y}_{ih} \\
 &= \sum_{s=1}^2 \sum_{h=1}^{n_i} {}_sM_i {}_sY_{ih} \quad (3.1)
 \end{aligned}$$

여기에서 s 는 성별(남-여)을 나타내는 첨자, n_i 는 경제활동인구조사에서 i 번째 지역의 표본조사구 수, ${}_sY_{ih}$ 는 각 성별에 대해서 i 번째 지역의 표본조사구에서 조사한 실업자 수를 나타낸다. 승수 ${}_sM_i = {}_s\hat{X}_{i\cdot} / {}_sX_{i\cdot}$ 은 $\hat{Y}_{i\cdot}$ 이 불편추정량이 되도록 산정한다. 여기에서 ${}_s\hat{X}_{i\cdot}$ 은 i 번째 지역에 대한 15세 이상의 상주 추계인구를 나타내며, ${}_sX_{i\cdot}$ 는 경제활동인구조사에서 조사된 15세 이상의 조사인구를 나타낸다.

직접추정량 $\hat{Y}_{i\cdot}$ 의 분산은 다음 (3.2)식과 같이 주어진다.

$$\begin{aligned}
 Var(\hat{Y}_{i\cdot}) &= \sum_{s=1}^2 Var({}_s\hat{Y}_{i\cdot}) + 2Cov({}_1\hat{Y}_{i\cdot}, {}_2\hat{Y}_{i\cdot}) \quad , \quad i=1,2,\dots,I \\
 &= \sum_{s=1}^2 {}_sM_i^2 Var(\sum_{h=1}^{n_i} {}_sY_{ih}) + 2{}_1M_i {}_2M_i Cov(\sum_{h=1}^{n_i} {}_1Y_{ih}, \sum_{h=1}^{n_i} {}_2Y_{ih}) \quad (3.2)
 \end{aligned}$$

$\hat{Y}_{i\cdot}$ 의 분산에 대한 추정은 다음 (3.3)식의 추정공식을 이용하여 계산된다.

$$\hat{Var}(\hat{Y}_{i\cdot}) = \sum_{s=1}^2 {}_sM_i^2 (\zeta_i \sum_{h=1}^{n_i} {}_sU_{ih}^2) + 2{}_1M_i {}_2M_i (\zeta_i \sum_{h=1}^{n_i} {}_1U_{ih} {}_2U_{ih}) \quad (3.3)$$

여기에서 ${}_sU_{ih} = d_s Y_{ih} - {}_s\rho_i \cdot d_s X_{ih}$, $d_s Y_{ih} = {}_sY_{ih} - {}_sY_{i,h+1}$,

$d_s X_{ih} = {}_sX_{ih} - {}_sX_{i,h+1}$, ${}_s\rho_i = {}_sY_{i\cdot} / {}_sX_{i\cdot}$,

$\zeta_i = [1 - n_i / (10 N_i)] n_i / [2(n_i - 1)]$ 이고, N_i 는 소지역 i 에 대한 모

집단의 조사구 수를 나타낸다

(2) 합성 추정량

대영역을 I 개의 시군구 단위의 소지역들로 분할하고, 대영역을 특성 기준에 따라 유사성을 갖는 J 개의 성별-연령대별 범주들로 구분할 때 i 번째 소지역의 합성추정량 $\hat{Y}_{i.}^S$ 는 다음과 같이 주어질 수 있다.

$$\hat{Y}_{i.}^S = \sum_{j=1}^J \eta_{ij} \Psi^{a.j.}, \quad i=1,2,\dots,I, \quad (3.4)$$

식(3.4)에서 가중치 $\eta_{ij} = (\xi_{it}^C / \xi_{it}^R) \xi_{ij} x_j$ 는 i 번째 소지역에서 j 번째 범주에 대한 경제활동 추정인구를 나타낸다. 여기에서 ξ_{it}^C 는 t 번째 해의 상주추정인구, ξ_{it}^R 는 같은 해의 주민등록인구, x_j 는 경제활동인구조사에서 각 성별에 대한 j 번째 범주의 경제활동 참가율을 나타낸다. $\Psi^{a.j.} = \hat{Y}_{.j.} / \sum_{i=1}^I \phi_{ij}$ ($j=1,2,\dots,J$)는 경제활동인구조사에서 추정된 j 번째 범주에 대한 실업률을 나타낸다.

소지역 i 에서 j 번째 범주에 대한 경제활동인구 η_{ij} 를 상수로 가정한다면 합성추정량 $\hat{Y}_{i.}^S$ 의 분산은 다음 (3.5)식과 같이 주어질 수 있다.

$$\text{Var}(\hat{Y}_{i.}^S) = \sum_{j=1}^J \eta_{ij}^2 \text{Var}(\Psi^{a.j.}) + 2 \sum_{j < l} \eta_{ij} \eta_{il} \text{Cov}(\Psi^{a.j.}, \Psi^{a.l.}), \quad i=1,2,\dots,I \quad (3.5)$$

합성추정량 $\hat{Y}_{i.}^S$ 의 추정분산은 다음 (3.6)식으로부터 계산될 수 있다.

$$\begin{aligned} \hat{\text{Var}}(\hat{Y}_{i.}^S) &= \sum_{j=1}^J \eta_{ij}^2 \left(\frac{1}{\sum_{i=1}^I \phi_{ij}} \right)^2 \hat{\text{Var}}(\hat{Y}_{.j.}) \\ &\quad + 2 \sum_{j < l} \eta_{ij} \eta_{il} \left(\frac{1}{\sum_{i=1}^I \phi_{ij}} \right) \left(\frac{1}{\sum_{i=1}^I \phi_{il}} \right) \hat{\text{Cov}}(\hat{Y}_{.j.}, \hat{Y}_{.l.}) \\ &= \sum_{j=1}^J \eta_{ij}^2 \left(\frac{1}{\sum_{i=1}^I \phi_{ij}} \right)^2 \left(M_j^2 \zeta_j \sum_{i=1}^I \sum_{h=1}^{n_i} U_{ih}^2 \right) \end{aligned}$$

$$+ 2 \sum_{j=1}^I \eta_{ij} \eta_{il} \left(\frac{1}{\sum_{i=1}^I \psi_{ij}} \right) \left(\frac{1}{\sum_{i=1}^I \psi_{il}} \right) \left(M_j M_l \zeta_j \sum_{i=1}^I \sum_{h=1}^{n_i} U_{ijh} U_{ilh} \right), (3.6)$$

여기에서 $U_{ijh} = d_j Y_{ijh} - \rho_j \cdot d_j X_{ijh}$, $d_j Y_{ijh} = Y_{ijh} - Y_{ij, h+1}$,

$$d_j X_{ijh} = X_{ijh} - X_{ij, h+1}, \quad \rho_j = Y_{\cdot j} / X_{\cdot j},$$

$\zeta_j = [1 - n_j / (10N_j)] n_j / [2(n_j - 1)]$. n_j 는 경제활동인구조사에서 j 번째

범주에 대한 표본 조사구 수를 나타내며, N_j 는 j 범주에 대한 모집단의 조사구 수를 나타낸다.

(3) 복합 추정량

i 번째 소지역에 대한 복합추정량 $\hat{Y}_{i \cdot}^C$ 는 다음 식을 이용하여 추정될 수 있다.

$$\hat{Y}_{i \cdot}^C = \hat{\omega}_{i(opt)} \hat{Y}_{i \cdot} + (1 - \hat{\omega}_{i(opt)}) \hat{Y}_{i \cdot}^S, \quad i=1, 2, \dots, I$$

여기서 가중값 $\hat{\omega}_{i(opt)} = \frac{\hat{V}ar(\hat{Y}_{i \cdot}^S)}{\hat{V}ar(\hat{Y}_{i \cdot}^S) + \hat{V}ar(\hat{Y}_{i \cdot})}$, $i=1, 2, \dots, I$ 이다.

직접추정량과 합성추정량의 공분산 $Cov(\hat{Y}_{i \cdot}, \hat{Y}_{i \cdot}^S) = 0$ 의 가정 하에 복합추정량의 분산 추정은 다음 식으로부터 계산될 수 있다.

$$\hat{V}ar(\hat{Y}_{i \cdot}^C) = \hat{\omega}_{i(opt)}^2 \hat{V}ar(\hat{Y}_{i \cdot}) + (1 - \hat{\omega}_{i(opt)})^2 \hat{V}ar(\hat{Y}_{i \cdot}^S)$$

광주광역시와 충청북도의 조사된 자료에서 계산한 성별-연령대별 실업률과 경제활동인구에 대한 요약이 다음 표와 같다.

34세 이하에서는 남자의 실업률이 7.2%이고 여자는 5.99%이며 35세 이상에서는 남자와 여자가 공히 2.9%로 유사하다.

<표 3.4> 광주광역시의 경제활동인구와 실업률

구 분	연령	경제활동인구		실업률	
		남자	여자	남자	여자
5월	15-34	106,077	101,708	0.0720	0.0599
	35+	256,0236	149,288	0.0297	0.0292
6월	15-34	114,226	113,559	0.0717	0.0836
	35+	208,131	155,443	0.0268	0.0344

충북의 실업률은 성별로도 차이가 있지만 시지역과 읍면지역 간에도 차이가 크다. 34세 이하의 남자의 경우 시지역은 6.17%이고 읍면지역은 4.23%이다.

<표 3.5> 충북 경제활동인구와 실업률

구분	그 룹	연령	경제활동인구		실업률	
			남자	여자	남자	여자
5월	시지역	15-34	63,048	57,578	0.0617	0.0503
		35+	152,951	102,481	0.0382	0.0106
	군지역	15-34	37,377	22,943	0.0423	0.0440
		35+	132,399	116,479	0.0139	0.0065
6월	시지역	15-34	65,473	56,778	0.0647	0.0510
		35+	154,438	98,367	0.0349	0.0221
	군지역	15-34	36,824	21,505	0.0357	0.0349
		35+	131,776	115,525	0.0120	0.0022

< 표 3.6 >은 기존 조사구에서 구별 실업자 추정값과 추정오차가 주어졌으며, <표 3.7>에는 증편한 조사구를 근거로 산정한 구별 실업자의 추정값과 추정오차이다.

<표 3.6> 설계기반 소지역 추정값과 상대표준오차(2001년 5월)

시 도	시군구	직접추정법			합성추정법			복합추정법		
		$\hat{Y}_{i.}$	S.E	RSE	$\hat{Y}_{i.}^s$	S.E	RSE	$\hat{Y}_{i.}^c$	S.E	RSE
광 주 광 역 시	동구	2,807	976	34.77	2,285	281	10.65	2,397	270	10.18
	서구	4,837	1,164	24.06	5,168	627	10.50	5,165	552	9.66
	남구	5,289	1,137	21.50	4,253	518	10.54	4,499	471	9.46
	북구	8,629	1,256	14.56	8,602	1,046	10.53	8,498	804	8.55
	광산구	2,761	830	30.06	4,015	482	10.39	3,764	417	10.01
	합계	24,323			24,323			24,323		

직접 추정값의 조사구 추가에 따른 상대표준오차의 향상은 3.3%에서 7.1%로 크지 않은 것으로 나타나고 있다.

<표 3.7> 설계기반 추정값과 상대표준오차(2001년 5월)

시 도	시군구	직접추정법			합성추정법			복합추정법		
		$\hat{Y}_{i.}$	S.E	RSE	$\hat{Y}_{i.}^s$	S.E	RSE	$\hat{Y}_{i.}^c$	S.E	RSE
광 주 광 역 시	동구	3,403	968	27.37	2,293	276	9.98	2,505	266	9.42
	서구	4,922	1,188	23.23	5,160	617	9.91	5,312	548	9.15
	남구	5,818	1,100	18.19	4,259	509	9.91	4,699	462	8.72
	북구	7,180	1,086	14.55	8,588	1,029	9.93	7,973	747	8.31
	광산구	3,000	717	23.00	4,023	474	9.77	3,834	396	9.16
	합계	24,323			24,323			24,323		

< 표 3.8 >은 조사구 증편에 따른 상대표준오차의 이득을 정리한 것이다.

<표 3.8> 설계기반 추정량들의 상대효율 이득(2001년 5월)

시도	시군구	직접추정량			합성추정량			복합추정량		
		RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)	RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)	RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)
광주광역시	동구	34.77	27.37	21.28	10.65	9.98	6.29	10.18	9.42	7.47
	서구	24.06	23.23	3.45	10.50	9.91	5.62	9.66	9.15	5.28
	남구	21.50	18.19	15.40	10.54	9.91	5.98	9.46	8.72	7.82
	북구	14.56	14.55	0.07	10.53	9.93	5.70	8.55	8.31	2.81
	광산구	30.06	23.00	23.49	10.39	9.77	5.97	10.01	9.16	8.49
	평균			12.74			5.91			6.37

< 표 3.9 >는 기존 조사구 자료에서 시군별로 추계한 실업자의 추정값과 추정오차 및 상대표준오차를 정리한 것이다.

<표 3.9> 설계기반 소지역 추정값과 상대표준오차(2001년 5월)

시도	시군구	직접추정법			합성추정법			복합추정법		
		$\hat{Y}_{i.}$	S.E	RSE	$\hat{Y}_{i.}^s$	S.E	RSE	$\hat{Y}_{i.}^c$	S.E	RSE
충청북도	청주시	9,929	1,948	19.62	9,341	1,248	11.51	9,434	1,051	9.94
	충주시	3,296	820	24.88	2,837	393	11.93	2,939	354	10.74
	제천시	751	475	63.25	1,798	250	11.97	1,604	221	12.29
	보은군	263	227	86.31	300	89	29.28	322	83	27.76
	옥천군	790	333	42.15	576	178	30.53	678	157	24.96
	영동군	252	189	75.00	577	178	30.48	461	130	30.37
	괴산군	779	649	83.31	295	88	29.43	331	87	28.34
	음성군	1,294	560	43.28	882	262	29.34	1,040	237	24.56
	청원군	767	307	40.03	1,343	413	30.39	1,054	246	25.15
	진천군	779	-	-	610	184	29.82	665	184	29.82
	단양군	0	-	-	341	102	29.57	372	102	29.53
합계	18,900			18,900			18,900			

< 표 3.10 >은 충북의 증편된 조사구를 근거로 시군별 실업자를 산정한 결과이다. <표 3.9>와 비교하였을 때 제천시의 직접 추정값의 변동이 매우 크게 나타났으나 복합 추정값에서는 추정값의 차이가 크지 않을 뿐 아니라 상대표준오차도 크게 향상되지 않음을 보이고 있다. 여기서 제천시의 직접 추정값의 변동이 큰 이유는 조사구가 4개에서 14개로 증가되어 적은 표본조사구에서 산정한 직접 추정량이 불안정함을 보이기 때문일 것이다.

< 표 3.10 > 설계기반 추정값과 상대표준오차(2001년 5월)

시 도	시군구	직접추정법			합성추정법			복합추정법		
		\hat{Y}_i	S.E	RSE	\hat{Y}_i^s	S.E	RSE	\hat{Y}_i^c	S.E	RSE
충 청 북 도	청주시	8,357	1,449	19.58	9,351	1,083	10.42	9,036	867	9.30
	충주시	3,084	731	26.77	2,831	338	10.74	2,980	307	9.99
	제천시	2,535	598	26.64	1,794	213	10.68	1,960	201	9.94
	보은군	749	183	23.05	307	44	15.55	341	43	13.83
	옥천군	837	224	25.25	568	85	16.19	625	80	14.04
	영동군	250	148	55.85	567	86	16.41	504	74	16.12
	괴산군	503	349	65.48	301	43	15.47	309	43	15.30
	음성군	661	329	46.93	898	128	15.42	893	119	14.64
	청원군	918	319	32.79	1,324	198	16.19	1,267	168	14.56
	진천군	584	300	48.47	613	89	15.70	627	85	14.89
	단양군	422	208	46.53	346	50	15.63	358	48	14.72
	합계	18,900			18,900			18,900		

< 표 3.11 >에서는 기존 조사구와 증편된 조사구의 추정값의 상대표준오차들의 이득을 산정한 결과이다. 소지역 추정법을 적용하더라도 시·군의 조사구는 최소한 4개 이상은 되어야 CV가 25%이하가 됨을 알 수 있다.

< 표 3.11 > 설계기반 추정량들의 상대효율 이득(2001년 5월)

시도	시군구	직접추정량			합성추정량			복합추정량		
		RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)	RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)	RSE (기존 조사)	RSE (증편 조사)	효율 이득 (%)
충청북도	청주시	19.62	19.58	0.20	11.51	10.42	9.47	9.94	9.30	6.43
	충주시	24.88	26.77	-7.60	11.93	10.74	9.97	10.74	9.99	6.98
	제천시	63.25	26.64	57.88	11.97	10.68	10.78	12.29	9.94	19.12
	보은군	86.31	23.05	73.29	29.28	15.55	46.89	27.76	13.83	50.18
	옥천군	42.15	25.25	40.09	30.53	16.19	46.97	24.96	14.04	43.75
	영동군	75.00	55.85	25.53	30.48	16.41	46.16	30.37	16.12	46.92
	괴산군	83.31	65.48	21.40	29.43	15.47	47.43	28.34	15.30	46.01
	음성군	43.28	46.93	-8.43	29.34	15.42	47.44	24.56	14.64	40.39
	청원군	40.03	32.79	18.09	30.39	16.19	46.73	25.15	14.56	42.11
	진천군	-	48.47	-	29.82	15.70	47.35	29.82	14.89	50.07
	단양군	-	46.53	-	29.57	15.63	47.14	29.57	14.72	50.22
평균			24.49			36.94			36.56	

◎ < 표 3.6 > ~ < 표 3.11 >에서 소지역 추정법의 효과와 표본 조사구의 증편에 대한 효과를 다음과 같이 요약할 수 있다.

☆ 표본 조사구를 다수 늘리더라도 직접 추정값의 효율은 크게 향상되지 않는다.

☆ 표본 조사구의 증편 후에 직접 추정값들의 효율 이득보다 기존의 경제활동인구조사에서 합성 추정값 또는 복합 추정값들의 효율 이득이 훨씬 높게 나타났다.

4. 시군구 실업자 추정 과제

4.1 효율적인 표본설계

소지역 추정법을 적용하여 시군구의 실업자 추정을 효율적으로 할 수 있는 표본설계의 연구가 최우선 과제임

- 모집단 구조분석

성별-연령대별로 유사한 시군구를 묶어서 총화하는데 필요한 정보 파악

시·도 경계를 초월하고, 산업적 특성과 지리적인 인접성을 고려하여 인근 시·도를 하나의 집락으로 묶는 분할방안

10%표본 조사구의 자료를 분석하여 실업률을 설명할 수 있는 보조 변수와 실업률의 지리적인 특성을 파악

- 통계생산 단위 결정 : 시군구 중에서 인구수 또는 정책적으로 중요도에 따라서 매월 실업통계를 생산해야할 시군구를 선정하고 소규모의 시군구에 대해서는 분기별로 실업통계를 발표하더라도 정책평가 또는 정책수립에 유익한 정보로 활용 가능(CV가 20-25%가 될 수있도록 단위 선정)

- 세분화된 가중값 체계 연구 : 50개 세분화된 총화로 50종의 가중값 설정하고 있으나 연령대별, 직업별 또는 학력별로 다양한 가중값을 부여하여 정확한 추정값을 산정토록 함.

- 표본 조사구의 크기와 각 층별로 표본 배분법

조사인력과 목표정도의 크기에 따라 적정한 표본크기 산정

시군구에 최소한 표본 조사구를 배분 후에 남은 조사구는 최적 배분법으로 배정(Top-down, Bottom-up의 절충)

- 표본 관리법과 조사방법의 개선 연구

일률적인 방문면접조사의 효율성을 검토 후에 전화조사 또는 인터넷조사를 병행하는 방안검토(광역시, 시지역과 농촌지역 구분)

4.2 소지역 추정법 연구

- 모형기반 소지역 추정법 개발 연구
- 행정정보조정보 이용방안 연구 : 2000년 인구주택총조사 자료, 주민등록 인구자료, 노동부의 구직등록데이터베이스와 실업보험청구자료.
- 시계열 모형 이용 추정법 연구 : 연속조사에서 과거 조사정보를 이용
- 합성 추정법과 복합 추정법에서 평균제곱오차의 추정 연구

5. 결 언

시군구 실업통계를 생산할 수 있는 완벽한 소지역 추정법의 연구와 적용도 중요하지만 현재의 문제를 개선할 수 있는 소지역 추정량을 우선 적용한 후에 점진적으로 개선 발전시키도록 지속적인 연구가 더 중요함.

모든 시군구의 실업통계를 생산하는 것보다 일정규모(인구 10만)이상의 시군의 실업통계를 우선적으로 소지역 추정법으로 추계하여 발표하는 방안

통계청과 대학간의 협동연구체제를 구성하여 통계의 신뢰성과 활용성을 제고하고 대학에서는 실용적인 교육을 추진할 수 있음.

21세기에는 통계분야에서도 선진국으로 발돋움할 수 있는 계기를 소지역 추정법 연구의 활성화를 통해서 마련할 수 있을 것임.