

소지역 추정법에 관한 비교연구¹

박종태², 이상은³

요약

직접(direct) 추정법, 합성(synthetic) 추정법, 복합(composite) 추정법, 베이스(Bayes) 추정법 등 소지역 추정법들의 효율성을 비교, 분석하고자 '98 경제활동 인구조사에서 경기도의 실제 자료를 이용하여 각 시부지역의 실업자수 추정값의 편의(bias)와 평균제곱 오차(MSE)를 모의실험을 통해 계산하였다.

주제어: 베이스 추정, 복합 추정, 소지역 추정, 직접 추정, 합성 추정.

1. 서론

소지역 추정은 표본의 크기가 작은 지역(small area)이나 나이, 연령, 교육정도, 소득수준 등과 같은 변수의 특정 분류된 소영역(small domain)에 대한 다양한 통계를 생산하는데 이용되고 있는 추정 방법이다. 실제로 미국, 캐나다, 이탈리아, 호주 등의 선진국가에서는 학자들과 정부기관과의 공동 협조로 소지역 추정 연구에 상당한 이론적 결과들을 축적해 왔으며, 정부 차원의 통계조사인 인구(센서스후의 인구수), 소득(4인 가구의 소득 중앙값), 노동력(고용률, 실업률), 농업(작물의 재배면적), 보건(유아와 산모의 건강상태, 치과방문자수) 통계 등의 다양한 분야에서 이를 활용하고 있다.

그러나 국내의 경우 전국적인 대규모의 표본조사와 관련된 표본설계가 대부분 전국을 16개 시도로 층화한 대지역 단위의 통계생산을 목적으로 하였기 때문에 시, 군부내의 작은 지역 단위의 통계가 필요할 경우에는 직접 조사된 자료만을 근거로 신뢰성 있는 통계생산이 불가능하므로 행정정보고 자료나 센서스 자료의 특성을 이용한 소지역 추정 방법에 대한 연구가 선행되어야 할 것이다. 뿐만 아니라 우리나라도 지방자치제 실시 이후로 지방자치단체 자체의 발전과 주민의 삶의 질과 아울러 알 권리를 충분히 만족시켜 주기 위한 정책 수립에 지역통계의 절실한 필요성이 대두되기 시작한 만큼, 정확한 지역 통계는 각 지역의 균형 있는 발전과 중앙 정부의 합리적인 예산 배분을 가져올 것이다. 소지역 추정, 특히 실업자수와 관련된 연구를 소개하면 Falorsi et al.(1994)은 이태리 노동력조사에 대하여 소지역 추정법들을 비교하였다. 또한 이상은 등(1996)은 1994년 경제활동인구조사 자료를 이용하여 두 지역에 대한 실업률의 추정을 위한 직접, 합성, 복합 추정법 등의 3가지 추정값의 안정성을 변동계수(Coefficient of Variation; CV)로 비교하였다.

이 논문에서는 소지역과 아울러 상관있는 지역에 대해 매우 제한된 정보만을 가지고 소지역의 실업자수를 추정하는 데 연구의 초점을 두고 있다. 이와 같은 취지로 본 연구에서

¹ 이 논문은 1999년도 한국과학재단의 해외연수지원에 의해 연구되었음.

² 경기도 평택시 용이동 평택대학교 정보과학과 조교수

³ 경기도 수원시 팔달구 이의동 경기대학교 응용정보통계학과 조교수

는 국내 노동력 통계조사와 관련된 경제활동 인구조사에 대하여 대표적인 소지역 추정법으로 알려진 직접, 합성, 복합, 베이스에 기초한 추정법 등을 비교하여 보고자 한다. 본 논문의 2절에서는 실업통계 추정을 위한 소지역 추정법의 이론적인 내용을 살펴보고 3절에서는 1998년도 경제활동 인구조사의 경기도 각 시부 지역을 모집단으로 하여 소지역에서의 실업자수에 대한 각 추정값의 편의(bias)와 평균제곱오차(mean square error; MSE)를 통계적 모의실험을 통해 계산한다. 마지막으로 각 소지역 추정법의 안정성을 모의실험 결과를 토대로 비교, 분석한다.

2. 소지역 추정법

표본조사 자료를 이용하여 특정집단의 총계나 평균을 추정하는 문제는 조사연구자들에게 일상적인 것이다. 특정집단이 크고 표본크기가 충분할 때 조사자료 자체에 의한 추정값이나 조사설계에 의한 추정값은 상당히 정도가 높다. 그러나, 특정집단이 작고 표본의 크기가 작을 때 조사자료 자체에 의한 추정량은 변동량이 커서 추정값으로서 이용하기에는 부적절하다. 이러한 문제를 해결하기 위해서 소지역 추정기법의 연구가 시작되었다.

만약 특정 소지역에 대한 통계를 생산하기 위한 표본설계시 종전의 조사설계법 위주로만 표본을 설계할 경우, 그에 따른 추정값에서 발생할 수 있는 2가지의 문제점을 생각할 수 있다. 첫째, 어떠한 표본에 포함된 특정 소지역에 대한 추정값 자체는 긍정적이라 할 수 있다. 하지만 작은 표본 규모로 인한 큰 표본오차 때문에 추정값에 대한 신뢰도가 떨어져 통계적 의미를 잃게 된다. 둘째, 어떠한 집단내에서 특이한 집단에 대한 표본수는 당연히 작을 것이다. 따라서 특정집단의 소지역 추정값을 조사설계법에 의해서만 표본설계를 한다면, 작은 표본규모로 인해 추정값이 0의 값을 가질 수 있다. 이는 자료 자체뿐만 아니라 추정값으로서의 통계적 의미도 없는 것이다. 또한 실제 적용시 잘못된 판단을 유도할 수 있다.

따라서 이러한 문제를 해결하기 위해서 소지역 자료와 상관관계가 있는 인구센서스(census) 자료나 행정자료 등을 보조정보로 이용한 모형식을 통해서 추정하게 된다. 이와 같은 관점에서 소지역 추정에 대한 다양한 베이스 방법들이 Ghosh와 Lahiri(1987), Ghosh, Datta와 Fay(1991) 등에 의해 연구되어 왔다.

2.1 직접 추정법(Direct estimation)

소지역에 대한 직접 추정은 조사된 자료 그 자체만을 이용하거나 또는 가능한 경우 그 지역의 행정자료나 센서스 자료를 보조정보로 이용하기도 한다.

소지역 a 에서 변수 Y 의 총계에 대한 가장 단순한 직접 추정량은

$$\hat{Y}_{a,D} = \sum_{i \in s_a} w_i y_i \quad (1)$$

여기서 s_a 는 소지역 a 에서 추출된 표본이며 w_i 는 i -번째 관측값의 가중값으로 이는 센서스나 행정자료에서 구해진다. $\hat{Y}_{a,D}$ 는 불편추정량이지만 소지역내에서의 표본크기가 랜덤하기 때문에 변동량이 크다. 즉, 전체에 대한 표본은 고정표본이지만 소지역 a 에 대한 표본은 랜덤이므로 상당히 작을 수도 있으며 전혀 개체가 없을 수도 있다.

만약 소지역 a 의 모집단의 크기 N_a 가 이용 가능한 경우 다음과 같은 사후 추정량을 이용한다.

$$\hat{Y}_{a,P} = N_a \frac{\sum_{i \in s_a} w_i y_i}{\hat{N}_a} \quad (2)$$

여기서 $\hat{N}_a = \sum_{i \in s_a} w_i$ 이다. 식 (2)의 $\hat{Y}_{a,P}$ 는 식 (1)의 $\hat{Y}_{a,D}$ 보다 변동량은 작으나 비추정에 의한 편의(bias)의 가능성이 있다.

2.2 합성 추정법(Synthetic estimation)

간접 추정(indirect estimation)의 한 방법으로서 합성추정을 Gonzalez(1973)은 “소규모의 지역은 대규모의 지역과 같은 특성을 가진다는 가정하에서 대규모의 지역으로부터 얻어진 불편추정값을 소규모 지역에 대한 추정값으로 보며, 이 추정값을 합성추정값이라고 한다.”라고 정의했다.

합성추정법은 소지역 표본 추정값 중 고전적이고 가장 흔히 쓰이는 방법이다. 이 방법은 추정 기법이 간단하고 원 표본설계의 응용이 가능하며 추정값을 구하려는 소지역과 흡사한 다른 소지역의 정보를 빌려서 추정값의 정도를 높이는 방법이다.

대지역을 A 개 소지역으로 분할하고 대지역에서 특성기준에 따른 범주를 G 개로 분류한다고 가정하자. 이 때 g 범주의 대지역에 대한 총계 Y_g 는 $Y_g = \sum_a Y_{ag}$ 가 된다. 여기서 Y_{ag} 는 소지역 a 에서 g 범주의 관측값이다. 또한 보조변수 X 에 대한 정보가 이용가능할 때 소지역 a 에 대한 합성추정량은 다음과 같다.

$$\hat{Y}_{a,S} = \sum_g \frac{X_{ag}}{X_{.g}} \hat{Y}_{.g}. \quad (3)$$

여기서 X_{ag} 는 소지역 a 에서 g 범주에 대한 보조정보이고, $X_{.g}$ 는 g 범주의 대지역에 대한 보조변수 X 의 총계이다. $\hat{Y}_{.g}$ 는 $\hat{Y}_{.g} = \frac{y_{.g}}{x_{.g}} X_{.g}$ 로 정의된 비추정량이며, $y_{.g}$ 는 g 범주에 대한 변수 Y 의 표본총계이고 $x_{.g}$ 는 g 범주에 대한 보조변수 X 의 표본총계이다.

2.3 복합 추정법(Composite estimation)

앞에서 언급한 직접 추정량은 편의가 작은 대신 변동량은 크지만 합성 추정량은 그 반대로 변동량이 작은 대신에 큰 편의를 갖는다. 따라서 직접 추정량 이용시의 큰 변동량과 합성 추정량 이용시에 발생하는 편의의 크기를 동시에 줄여주는 추정방법이 복합 추정법이다. 복합 추정량은 직접 추정량과 합성 추정량을 선형결합하는 것으로 일반적으로 다음과 같이 정의된다.

$$\hat{Y}_{a,C} = \alpha \hat{Y}_{a,D} + (1 - \alpha) \hat{Y}_{a,S}. \quad (4)$$

여기서 α 는 $\hat{Y}_{a,C}$ 의 MSE를 최소화하는 가중값이며 다음 식과 같다.

$$\alpha = \frac{\text{MSE}(\hat{Y}_{a,S}) - E(\hat{Y}_{a,S} - Y_a)(\hat{Y}_{a,D} - Y_a)}{\text{MSE}(\hat{Y}_{a,S}) + \text{MSE}(\hat{Y}_{a,D}) - 2E(\hat{Y}_{a,S} - Y_a)(\hat{Y}_{a,D} - Y_a)}. \quad (5)$$

또한 식 (5)에서 공분산을 무시한 경우 α 의 최적값 α_{opt}^* 는 다음과 같이 근사화 할 수 있는데 이 값이 Schaible(1978)에 의해 정의된 가중값이다.

$$\alpha_{\text{opt}}^* = \frac{\text{MSE}(\hat{Y}_{a,S})}{\text{MSE}(\hat{Y}_{a,S}) + \text{MSE}(\hat{Y}_{a,D})}. \quad (6)$$

본 연구에서는 복합추정값을 계산하기 위해 식 (6)에 의한 α 의 최적값 α_{opt}^* 을 이용할 것이다.

2.4 베이즈 추정법(Bayes estimation)

소지역에 대한 관심있는 변수의 총계를 추정하기 위한 베이즈 방법을 소개하면 다음과 같다. 여기서 고려되어 지는 모형은 다음과 같이 주어진다. 소지역 a 의 i -번째 표본단위에서 관찰된 변수, 즉 실업자수를 y_{ai} 라고 하면 가구당 평균 실업자수 μ_a 가 주어질 때 이의 조건부 분포는 $y_{ai}|\mu \sim N(\mu_a, \sigma^2)$, $a = 1, 2, \dots, A$, $i = 1, 2, \dots, n_a$ 을 따른다고 가정하자. 또한 실업자수와 상관관계가 높은 보조정보(예를 들면 15세 이상 인구수)를 x_a 라고 할 때 평균 μ_a 의 사전분포로 $\mu_a \sim N(\alpha_a + \beta_a \bar{x}_a, \tau^2)$ 을 따른다고 가정하자. 여기서 σ^2, τ^2 은 알려진 모수라고 가정한다. 이 때 μ_a 의 사후분포는 베이즈 방법에 의해 쉽게 얻을 수 있고 이는 다음과 같이 주어진다.

$$\mu_a|y_a \sim N(\hat{\mu}_a, s^2), \quad \hat{\mu}_a = (1 - \lambda_a)\bar{y}_a + \lambda_a(\hat{\alpha}_a + \hat{\beta}_a \bar{x}_a). \quad (7)$$

여기서 $s^2 = \sigma^2 \tau^2 / (\sigma^2 + n_a \tau^2)$, $\bar{y}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} y_{ai}$, $\lambda_a = \sigma^2 / (\sigma^2 + n_a \tau^2)$ 이다. 또한 $\hat{\alpha}_a$ 와 $\hat{\beta}_a$ 는 각 소지역의 전체 가구에서 관찰된 실업자수 Y_{ai} 와 이와 밀접한 보조정보 X_{ai} 의 선형회귀모수의 추정값이다. 따라서 손실함수가 오차제곱함수로 주어질 때, 즉 $(\hat{\mu}_a, \mu_a) = (\hat{\mu}_a - \mu_a)^2$ 일 때 μ_a 의 베이즈 추정량 $\hat{\mu}_{a,B}$ 는 손실함수의 기대값을 최소화하는 $\hat{\mu}_{a,B} = E(\mu_a|y_a)$ 이 된다. 즉 μ_a 의 사후분포의 기대값이 된다. 그러므로 보조변수 X 에 대한 정보가 이용가능할 때 소지역 a 에 대한 실업자수 Y_a 의 베이즈 추정량은 단순한 비추정에 의해 다음과 같이 주어진다.

$$\hat{Y}_{a,B} = w_a n_a \hat{\mu}_{a,B}, \quad w_a = X_a / x_a. \quad (8)$$

여기서 X_a 는 소지역 a 의 모집단과 관련된 보조정보이고 x_a 는 소지역 a 의 표본과 관련된 보조정보이다.

3. 모의실험

1998년 경제활동인구조사 자료중 경기도내 시부지역의 15세 이상 인구와 실업자 인구 자료를 이용하여 소지역 추정법을 비교하기 위한 모의실험을 수행하는데 각 추정값의 안정성을 검토하기 위하여 편의와 평균제곱오차를 계산하여 비교, 분석한다. 모의실험 방법은 다음과 같다: 경기도내 조사구 수가 2개 이상인 18개 시를 소지역으로 하고 1998년 중 실업률이 가장 높았던(9.3%) 12월의 표본조사 가구에서 관찰된 실제 자료를 모집단으로 가정하여 각 소지역에서 10%의 가구를 표본으로 추출하고 이를 1000회 반복 실험하였다. 가정된 모집단에 대한 실제 자료는 다음 표 1과 같다.

표 1 경기도의 경제활동인구조사 자료(1998. 12)

(단위: 개, 명)

구분	조사 구수	가구수	15세 이상 인구			경제활동인구			실업자 인구		
			남자	여자	합계	남자	여자	합계	남자	여자	합계
시부	98	2,326	2,704	2,957	5,661	2,116	1,315	3,431	207	136	343
군부	20	486	588	609	1,197	424	238	662	26	7	33
전체	118	2,812	3,292	3,566	6,858	2,540	1,553	4,093	233	143	376

각 소지역 추정량에 대해 실제 자료를 이용한 자세한 계산 과정은 다음과 같다: 첫째, 직접 추정량의 경우 소지역 a 에 대한 실업자수의 직접 추정값은 통계청(1996)에서 이용하고 있는 각 시의 동부와 읍면부에 대한 승수와 경제활동인구조사 실제자료를 이용하여 식 (1)에 의해 계산하였다; 둘째로 합성 추정량의 경우 소지역 a 에 대한 실업자수의 합성 추정

값은 식 (3)을 통해 계산하였다. 이 때 실업자와 밀접한 관련있는 경제활동인구를 보조정보로 이용하였고 범주는 성별(남/녀)로 하였다; 셋째, 복합 추정량의 경우 먼저 식 (4)에 있는 α 를 식 (6)을 통해 최적값 α_{opt}^* 로 계산하였고(표 3과 4에 제시) 이를 이용하여 최종적인 복합 추정값을 계산하였다; 마지막으로 베이즈 추정량의 경우는 우선 소지역 a 의 가구당 평균 실업자수의 베이즈 추정값 $\hat{\mu}_{a,B}$ 을 계산하기 위해 실제자료를 이용하여 σ^2 과 τ^2 을 구하였는데 이의 값은 $\sigma^2 = 0.166$, $\tau^2 = 0.006$ 이었다. 이 값을 통해 각 소지역 a 에서 λ_a 을 계산하였고(표 3과 4에 제시) 15세 이상 인구수를 보조정보로 이용하여 최종적으로 소지역 a 에 대한 실업자수 Y_a 의 베이즈 추정값을 식 (8)을 통해 계산하였다.

마지막으로 위에서 언급한 각 소지역 추정량을 비교하기 위한 편의와 평균제곱오차는 Falorsi et al.(1994)에서 제시한 것과 유사하게 다음과 같이 계산하였다.(결과는 표 3과 4에 제시)

$$BIAS_{a,m} = \frac{1}{R} \sum_{r=1}^R (\hat{Y}_{a,m}(r) - Y_a),$$

$$MSE_{a,m} = \frac{1}{R} \sum_{r=1}^R (\hat{Y}_{a,m}(r) - Y_a)^2.$$

여기서 $\hat{Y}_{a,m}(r)$ 은 소지역 a 의 추정법 m 에 의해 추정된 r -번째 반복실험에서의 실업자수이며, R 은 반복실험횟수이다. ($R = 1000$)

또한 소지역 전체에 대한 절대편의와 평균제곱오차의 평균은 다음과 같이 계산하였다.(결과는 표 2에 제시)

$$\overline{ABIAS}_m = \frac{1}{A} \sum_{a=1}^A |BIAS_{a,m}|,$$

$$\overline{MSE}_m = \frac{1}{A} \sum_{a=1}^A MSE_{a,m}.$$

여기서 A 는 전체 소지역의 수이다. ($A = 18$)

4. 결과 분석

우선 18개 소지역 전체에 대한 각 추정량의 절대편의와 평균제곱오차의 평균은 다음 표 2와 같다. 이 표를 보면 다음과 같은 결과를 알 수 있다:

(i) Bayes가 가장 작은 편의를 가진다. Synthetic의 경우가 다른 추정량보다 편의가 가장 크다. 또한 Composite의 편의가 Synthetic보다 낮게(대략 42%) 나타난다. Direct의 편의는 Bayes 만큼 무시할 만한 수준으로 작게 나타난다.

(ii) Bayes가 가장 작은 평균제곱오차를 가짐을 알 수 있다. 반면에 거의 비슷한 편의를 갖는 Direct의 경우는 MSE가 가장 크게 나타나기 때문에 효율적인 추정량이라 할 수 없다. 또한 Bayes의 평균제곱오차는 Synthetic이나 Composite 추정량보다는 상당히 낮다는 것을 알 수 있다.

표 2 추정량의 절대편의와 평균제곱오차의 평균

Estimator	ABIAS	MSE
Direct	0.349	213.225
Synthetic	7.099	88.405
Composite	5.007	57.960
Bayes	0.153	24.063

표 3 각 소지역에 대한 추정량의 편의

small area	Estimator				
	Direct	Synthetic	Composite	(α_{opt}^*)	Bayes (λ_a)
1	-0.134	17.704	12.405	(0.30)	-0.101 (0.51)
2	-0.225	5.425	5.022	(0.09)	-0.113 (0.63)
3	0.171	5.296	4.982	(0.06)	0.206 (0.76)
4	-0.156	18.048	7.815	(0.55)	-0.056 (0.61)
5	0.498	-5.580	-4.392	(0.25)	0.060 (0.61)
6	0.662	-3.056	-2.507	(0.19)	0.184 (0.64)
7	0.038	5.149	4.587	(0.12)	-0.046 (0.76)
8	-0.567	2.824	2.677	(0.05)	-0.238 (0.76)
9	-0.593	8.900	4.272	(0.49)	-0.274 (0.76)
10	0.332	7.278	6.157	(0.14)	0.200 (0.67)
11	0.077	5.411	4.120	(0.24)	0.099 (0.85)
12	0.029	-0.289	-0.301	(0.05)	-0.019 (0.85)
13	0.206	6.997	4.679	(0.34)	0.211 (0.80)
14	0.433	-11.963	-7.012	(0.45)	0.138 (0.83)
15	0.366	-1.614	-1.328	(0.10)	0.198 (0.70)
16	-0.656	2.239	2.066	(0.06)	-0.394 (0.72)
17	0.721	-11.983	-8.334	(0.28)	0.146 (0.74)
18	-0.412	8.031	7.478	(0.07)	-0.070 (0.67)

* Composite의 경우 괄호안의 수치는 α 의 최적값이며, Bayes의 경우는 λ_a 의 값임.

이제 소지역 각각에 대한 편의와 평균제곱오차를 비교해 보면 표 3, 4를 통해 알 수 있다.

(i) Synthetic과 Composite의 경우 몇몇 소지역(1, 4, 14, 17)에서는 편의가 아주 크게 나타났다. 그러나 MSE는 대체로 Direct 추정량보다 낮게 나타남을 알 수 있다. 18개의 소지역 중 9개 지역(1, 4, 9, 10, 11, 13, 14, 17, 18)에서는 Composite이 가장 효율적으로 나타났고, 4개 지역(5, 6, 12, 15)에서는 Synthetic이 다소 우세한 것으로 보였으며 나머지 5개 지역은 두 추정량이 MSE의 관점에서 비슷하게 나타났음을 알 수 있다.

(ii) Bayes의 경우 편의는 거의 모든 소지역에서 무시할 만한 수준을 보임을 알 수 있다. 뿐만 아니라 MSE도 5개 지역(5, 6, 12, 15, 16)을 제외한 모든 지역에서 Synthetic과 Composite 추정량보다 낮게 나타났음을 알 수 있다.

(iii) 이상의 결과에서 보는 바와 같이 편의와 평균제곱오차의 종합적인 측면을 고려해 볼 때 기존의 고전적인 세 방법(Direct, Synthetic, Composite)보다 Bayes 추정법이 가장 효율적이라는 것이 이 모의실험의 경험적 결과(empirical results)이다.

표 4 각 소지역에 대한 추정량의 평균제곱오차

small area	Estimator					
	Direct	Synthetic	Composite	(α_{opt}^*)	Bayes	(λ_a)
1	339.589	364.553	221.613	(0.30)	80.670	(0.51)
2	303.686	51.986	50.180	(0.09)	42.391	(0.63)
3	151.146	34.532	31.682	(0.06)	11.944	(0.76)
4	181.768	351.037	121.057	(0.55)	27.158	(0.61)
5	432.570	56.754	69.242	(0.25)	72.114	(0.61)
6	370.982	34.682	38.702	(0.19)	48.039	(0.64)
7	88.076	31.458	26.603	(0.12)	5.292	(0.76)
8	164.202	15.706	15.060	(0.05)	10.291	(0.76)
9	57.322	85.445	35.380	(0.49)	5.067	(0.76)
10	193.315	72.598	58.978	(0.14)	21.361	(0.67)
11	40.561	31.462	20.995	(0.24)	1.275	(0.85)
12	121.909	1.962	2.289	(0.05)	2.756	(0.85)
13	68.381	53.120	32.533	(0.34)	5.320	(0.80)
14	247.134	146.830	102.297	(0.45)	7.656	(0.83)
15	264.491	12.741	15.306	(0.10)	25.429	(0.70)
16	140.138	11.431	10.953	(0.06)	13.406	(0.72)
17	494.474	152.624	116.629	(0.28)	33.310	(0.74)
18	178.308	82.376	73.790	(0.07)	19.664	(0.67)

* Composite의 경우 괄호안의 수치는 α 의 최적값이며, Bayes의 경우는 λ_a 의 값임.

참고문헌

1. 이상은, 진영 (1996). 변동계수를 이용한 소지역 통계의 안정성 검토, 통계분석연구, 창간호, 23-40, 통계청.
2. Falorsi, P.D., Falorsi, S. and Russo, A. (1994). Empirical comparison of small area estimation methods for the Italian Labour Force Survey, *Survey Methodology*, Vol. 20, No. 2, 171-176.
3. Ghosh, M., Datta, G.S. and Fay, R.E. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation, In *Proceeding of the Bureau of the Census Annual Research Conference*, 63-79, Bureau of the Census, Washington, DC.
4. Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples, *Journal of the American Statistical Association*, Vol. 82, 1153-1162.

5. Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates, *Proceedings of the Social Statistics Section, American Statistical Association*, 33-36.
6. Schaible, W.L. (1978). Choosing weights for composite estimators for small area statistics, *Proceedings fo the Section on Survey Research Methods, American Statistical Association*, 741-746.

A Comparative Study of Small Area Estimation Methods ⁴

Jong Tae Park ⁵ and Sang Eun Lee ⁶

Abstract

Usually estimating the means is used for statistical inference. However depending the purpose of survey, sometimes totals will give the better and more meaningful in statistical inference than the means. Here in this study, we dealt with the unemployment population of small areas with using 4 different small area estimation methods: Direct, Synthetic, Composite, Bayes estimation. For all the estimates considered in this study, the average of absolute bias and mean square error were obtained in the Monte Carlo Study which was simulated using data from 1998 Economic Active Population Survey in Korea.

Key Words and Phrases: Bayes estimation, Compositie estimation, Direct estimation, small area estimation, Synthetic estimation.

⁴This research was supported by Korea Science and Engineering Foundation in 1999.

⁵Assistant Professor, Department of Information Science, Pyongtaek University, Pyongtaek, Kyonggi, 450-701, Korea

⁶Assistant Professor, Department of Applied Information Statistics, Kyonggi University, Suwon, Kyonggi, 442-760, Korea