

On Estimating the Odds Ratio between Male and Female Unemployment Rate in Small Area

Jong Tae Park¹⁾

Abstract

There are different kinds of methods to estimate the odds ratio for unemployment statistics in small areas, namely, the composite estimator, the Woolf estimator and the Mantel-Haenszel estimator. We can compare the reliability of these estimators according to the bias and MSE. The estimation procedures considered by this study have been applied to estimate the bias and MSE of the odds ratio between the male and female unemployment rate in some small areas. The Woolf estimator or the Mantel-Haenszel estimator is more stable than the composite estimator, but all these three estimators are similar to each other from the aspect of efficiency.

Keywords : Bias, Mean square error, Odds ratio, Small area

1. 서론

실업통계는 매월 정부주관의 노동력통계조사인 경제활동인구조사를 통하여 발표되고 있다. 그러나 이 실업통계에 있어 우리나라에서는 현재 광역시나 도 단위의 대지역에 대해서만 발표하고 있어 시군구단위의 소지역통계(small area statistics)가 없다는 것이 안타까운 현실이다. 특히, 실업률은 국가경제와 밀접하게 관련된 변수중의 하나이다. 따라서 미국, 일본 등을 비롯하여 각 국가들은 세계적인 장기간의 경기 불황으로 인한 실업률을 낮추기 위해 안간 힘을 쓰고 있고 우리나라도 예외는 아니다. 특히 우리의 경우 98년 외환위기 이후 광범위한 기업구조조정의 결과 해고, 명예퇴직 등의 사유로 인해 실업률이 외환위기 이전보다 상대적으로 높아졌고, 최근 들어서는 청년 실업의 증가로 인해 새 정부 들어 새로운 일자리를 창출하고자 노력함에도 불구하고 한번 높아진 실업률은 좀처럼 떨어지지 않고 있다. 이와 같은 국내외 사정 등으로 말미암아 외환위기 이전보다 실업통계에 대한 관심이 지대해 지고 있고, 더 나아

1) Associate Professor, Department of Digital Information & Statistics, Pyeongtaek University, Pyeongtaek, 450-701, Korea
E-mail : jtpark@ptu.ac.kr

가 지역자치단체들도 예전과 달리 소지역통계, 특히 경제관련 통계를 철저히 필요로 하고 있다.

이를 극복하기 위해 그동안 국내에서는 소지역의 실업통계에 대한 다양한 연구가 이루어져 왔다. 먼저, 이상은 등(1996)은 1994년 경제활동인구조사 자료를 이용하여 충청도의 두 소지역에 있어서 실업률에 대한 3가지 추정량들을 변동계수로 비교하였다. 또한 이상은(2001)은 소지역추정을 위해 베이즈 예측값을 이용하였다. 이계오(2000)는 인구통계학적 방법, 합성 추정, 복합 추정 등의 소지역 추정법을 이용하여 충북 시, 군, 구의 실업자 수를 추정하였다. 박종태 등(2001)은 직접추정, 합성추정, 복합추정, 베이즈 추정 등의 효율성을 비교하고자 모의실험을 통해 편의와 평균제곱오차를 계산하였다. 양영춘 등(2003)은 실업자 수의 소지역추정을 위해 Kalman Filter 기법을 이용하였고, 정연수 등(2003)은 시군구 실업자 총계 추정을 위해 표본설계에 기반을 둔 간접추정법을 제안하였다. 최근에 박종태 등(2003)은 경기지역의 시군단위 행정자치구역에 있어서 남녀별 실업률의 오즈와 오즈비를 추정하였고 이 추정치들을 소지역 별로 비교하였다.

본 논문에서는 경기도의 각 시군을 소지역으로 하여 각 지역에 있어서 남녀별 실업률의 오즈비(odds ratio)를 합동추정량, Woolf 추정량과 Mantel-Haenszel 추정량 등의 방법으로 추정하고 이들의 신뢰성을 편의(Bias)와 평균제곱오차(MSE)를 통해 검증하고자 한다. 2002년 12월 경기지역의 경제활동인구조사 자료로부터 24개 시군단위 행정자치구역들에 대한 남녀 그룹 간 실업률에 대한 오즈비를 추정한다. 또한 시군단위 오즈비 추정치들의 신뢰성은 이들 추정치들의 상대편의(RB: relative bias)와 상대평균 오차제곱근(RRMSE: relative root mean square error)에 의해 평가된다.

2. 소지역의 실업률에 대한 공통 오즈비의 추정

국내의 행정구역상 도나 광역시에 해당하는 어떤 대지역이 I 개의 시군구 등과 같은 소지역으로 분할되었다고 가정하자. 이 때 소지역 i 에 할당된 표본조사구수를 n_i 라 하고 이 중에서 k 번째 표본조사구에서 임의로 추출된 성별변수 X 와 경제활동변수 Y 의 값에 따라 다음과 같이 2×2 분할표로 분류되었다고 가정하자.

		Y		Total
		실업	취업	
X	1(남)	x_{k1}	$N_{k1} - x_{k1}$	N_{k1}
	2(여)	x_{k2}	$N_{k2} - x_{k2}$	N_{k2}

여기서 N_{k1} 과 N_{k2} 는 두 그룹의 표본크기이다. 이 때 남자, 여자 각각에 대한 성공(실업)의 오즈(odds)는 다음과 같이 주어진다.

$$odds_1 = \frac{\pi_1}{1 - \pi_1}, \quad odds_2 = \frac{\pi_2}{1 - \pi_2}.$$

여기서 π_1 은 남자에 대한 실업률이고 π_2 는 여자에 대한 실업률을 나타낸다.

소지역 i 의 k 번째 표본조사구에 대한 오즈비는 두 그룹의 성공확률, 즉 실업률의 비(ratio)로서 다음과 같이 주어진다.

$$\Theta_{ik} = \frac{odds_{k1}}{odds_{k2}} = \frac{\pi_{k1}/(1 - \pi_{k1})}{\pi_{k2}/(1 - \pi_{k2})}.$$

또한 표본오즈비는 두 그룹의 표본오즈들의 비로 다음과 같다.

$$\hat{\Theta}_{ik} = \frac{p_{k1}/(1 - p_{k1})}{p_{k2}/(1 - p_{k2})} = \frac{x_{k1}/(N_{k1} - x_{k1})}{x_{k2}/(N_{k2} - x_{k2})} = \frac{x_{k1}(N_{k2} - x_{k2})}{x_{k2}(N_{k1} - x_{k1})}.$$

여기서 p_{k1} 과 p_{k2} 는 k 번째 표본조사구에 대한 남녀 각각의 표본실업률이다.

이제 주어진 소지역 i 의 모든 표본조사구에서 남녀별 실업률에 대한 오즈비가 모두 동일하다고 가정한다. 즉, $\Theta_{i1} = \Theta_{i2} = \dots = \Theta_{ij} = \Theta_i$ 로 가정하자. 이 때 소지역 i 의 공통 오즈비인 Θ_i 를 추정하기 위해 일반적인 합동추정, Woolf 추정과 Mantel-Haenszel 추정 등을 고려하기로 한다.

2.1 합동추정량

소지역 i 에서의 공통 오즈비의 합동추정량은 일반적으로 다음과 같이 주어진다.

$$\hat{\theta}_i^p = \frac{x_{.1}(N_{.2} - x_{.2})}{x_{.2}(N_{.1} - x_{.1})},$$

여기서 $x_{.1} = \sum_{k=1}^{n_i} x_{k1}$, $x_{.2} = \sum_{k=1}^{n_i} x_{k2}$, $N_{.1} = \sum_{k=1}^{n_i} N_{k1}$, $N_{.2} = \sum_{k=1}^{n_i} N_{k2}$ 이다.

표본오즈비의 로그 변환 $\log \hat{\theta}_i^p$ 는 $\hat{\theta}_i^p$ 의 분포보다 더 정규분포에 가까운 표본추출분포를 갖고, 표본크기가 커지면 $\log \hat{\theta}_i^p$ 의 대표분포는 평균이 $\log \theta_i$ 이고 다음과 같은 근사적인 표준편차를 갖는 정규분포로 수렴한다고 알려져 있다.

$$\sigma(\log \hat{\theta}_i^p) \simeq \sqrt{\frac{1}{x_{.1}} + \frac{1}{N_{.1} - x_{.1}} + \frac{1}{x_{.2}} + \frac{1}{N_{.2} - x_{.2}}}.$$

그러므로 표본오즈비 $\hat{\theta}_i^p$ 의 대표분포는 근사적으로 로그 정규분포(log normal)에 수렴하게 된다. 따라서 소지역 i 에 대한 공통 오즈비의 합동추정량($\hat{\theta}_i^p$)의 편의(Bias)와 평균제곱오차(MSE)의 추정량은 각각 다음과 같이 주어진다

$$\begin{aligned}\widehat{Bias}(\hat{\theta}_i^p) &= \hat{\theta}_i^p \left[\exp\left(\frac{1}{2}\sigma^2(\log\hat{\theta}_i^p)\right) - 1 \right] \\ &\simeq \frac{1}{2}\hat{\theta}_i^p \left(\frac{1}{x_{.1}} + \frac{1}{N_{.1}-x_{.1}} + \frac{1}{x_{.2}} + \frac{1}{N_{.2}-x_{.2}} \right), \\ \widehat{MSE}(\hat{\theta}_i^p) &= (\hat{\theta}_i^p)^2 \left[\exp(2\sigma^2(\log\hat{\theta}_i^p)) - 2\exp\left(\frac{1}{2}\sigma^2(\log\hat{\theta}_i^p)\right) + 1 \right] \\ &\simeq (\hat{\theta}_i^p)^2 \left[\left(\frac{1}{x_{.1}} + \frac{1}{N_{.1}-x_{.1}} + \frac{1}{x_{.2}} + \frac{1}{N_{.2}-x_{.2}} \right) \right. \\ &\quad \left. + \frac{7}{4} \left(\frac{1}{x_{.1}} + \frac{1}{N_{.1}-x_{.1}} + \frac{1}{x_{.2}} + \frac{1}{N_{.2}-x_{.2}} \right)^2 \right].\end{aligned}$$

그러나 소지역 i 에 할당된 표본조사구 수가 충분하지 못하여 칸 도수가 아주 작거나 0인 경우에는 합동추정치 추정오차가 커져 편의와 MSE의 추정에 위의 식들을 사용하기가 어려울 것이다. 따라서 편의와 MSE 추정을 위한 하나의 대안으로써 잭나이프 추정방법을 고려하기로 한다. 잭나이프 추정에서는 서로 독립이고 동일한 임의 확률분포를 따르는 확률표본을 가정한다. 국내의 경제활동인구조사 표본은 2단 층화 추출로 추출되는 확률표본이다. 대지역내에서 표본조사구들이 일차적으로 계통추출된 후 각 표본조사구 내에서 일정량의 가구 수를 갖는 작은 구획들이 이차적으로 랜덤추출되어 구획 내의 가구단위들에 대해 조사가 이루어진다. 표본조사구내의 구획단위들은 서로 독립인 확률표본이고 이들의 관측값에 의해 표본조사구의 관측 값이 결정된다. 그러나 대지역내의 표본조사구들은 랜덤 수에 계통추출되므로 유사 독립표본으로 간주될 수는 있지만 정확히 독립인 확률표본들은 아니다.

대지역내의 표본조사구들을 서로 독립이고 동일한 임의 분포를 갖는 확률표본으로 가정한 상태에서 잭나이프 추정절차를 소개하면 다음과 같다. 잭나이프 추정에서 첫 번째 단계는 경제활동인구조사 자료로부터 반복표본을 생성하는 일이다. 우선 소지역 i 내에서 하나의 표본조사구가 교대로 선택되어 표본으로부터 제거된 후 나머지 표본조사구들에 대해서 합동추정치가 계산된다. 반복표본들은 표본조사구의 개수만큼 생성되며 이들 반복표본들을 이용하여 새로운 합동추정치들이 다시 계산된다.

소지역 i 에 있어서 남녀별 실업률의 오즈비에 대한 합동추정치들의 잭나이프 편의와 MSE 추정식은 각각 다음과 같이 주어진다.

$$\begin{aligned}\widehat{Bias}_J(\hat{\theta}_i^p) &= (n_i - 1) (\hat{\theta}_{i(\cdot)}^p - \hat{\theta}_i^p), \\ \widehat{Var}_J(\hat{\theta}_i^p) &= \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} [\hat{\theta}_{i(k)}^p - \hat{\theta}_{i(\cdot)}^p]^2, \\ \widehat{MSE}_J(\hat{\theta}_i^p) &= \widehat{Var}_J(\hat{\theta}_i^p) + [\widehat{Bias}_J(\hat{\theta}_i^p)]^2,\end{aligned}$$

여기서, $\widehat{\theta}_{i(k)}^p = \frac{(x_{.1} - x_{k1})(N_{.2} - N_{k2} - x_{.2} + x_{k2})}{(x_{.2} - x_{k2})(N_{.1} - N_{k1} - x_{.1} + x_{k1})}$, $\widehat{\theta}_{i(\cdot)}^p = \frac{1}{n_i} \sum_{k=1}^{n_i} \widehat{\theta}_{i(k)}^p$ 이다.

2.2 Woolf 추정량

Woolf(1955)에 의해 제안된 울프추정량은 근사적 분산의 역수를 가중치로 각각의 추정된 로그오즈비의 합으로 다음과 같이 주어진다.

$$\log \widehat{\theta}_i^W = \frac{\sum_{k=1}^{n_i} w_{ik} \log \widehat{\theta}_{ik}}{\sum_{k=1}^{n_i} w_{ik}},$$

여기서 $w_{ik}^{-1} = \sigma^2(\log \widehat{\theta}_{ik}) \simeq \frac{1}{x_{.1}} + \frac{1}{N_{.1} - x_{.1}} + \frac{1}{x_{.2}} + \frac{1}{N_{.2} - x_{.2}}$ 이다. 그리고 대표본 일 경우, $\log \widehat{\theta}_i^W$ 의 분산은 근사적으로 다음과 같이 주어진다.

$$\sigma^2(\log \widehat{\theta}_i^W) = \frac{1}{w_i},$$

여기서 $w_i = \sum_{i=1}^{n_i} w_{ik}$ 이다.

따라서 합동추정량의 경우와 마찬가지로 소지역 i 에 대한 공통 오즈비의 울프추정량 ($\widehat{\theta}_i^W$)의 편의와 평균제곱오차의 추정치는 근사적으로 다음과 같이 주어진다.

$$\begin{aligned} \widehat{Bias}(\widehat{\theta}_i^W) &= \widehat{\theta}_i^W [\exp(\frac{1}{2}\sigma^2(\log \widehat{\theta}_i^W)) - 1] \simeq \frac{\widehat{\theta}_i^W}{2w_i}, \\ \widehat{MSE}(\widehat{\theta}_i^W) &= (\widehat{\theta}_i^W)^2 [\exp(2\sigma^2(\log \widehat{\theta}_i^W)) - 2\exp(\frac{1}{2}\sigma^2(\log \widehat{\theta}_i^W)) + 1] \\ &\simeq (\widehat{\theta}_i^W)^2 (\frac{1}{w_i} + \frac{7}{4w_i^2}). \end{aligned}$$

Woolf 추정량의 편의와 MSE 추정에서도 합동추정량의 경우와 같이 잭나이프 추정 방법이 이용될 수 있으며 소지역 i 에서 오즈비에 대한 울프추정량의 편의와 MSE 추정치들은 다음 추정식으로부터 산정할 수 있다.

$$\widehat{Bias}_J(\widehat{\theta}_i^W) = (n_i - 1)(\widehat{\theta}_{i(\cdot)}^W - \widehat{\theta}_i^W),$$

여기서

$$\widehat{MSE}_J(\theta_i^{\widehat{W}}) = \widehat{Var}_J(\theta_i^{\widehat{W}}) + [\widehat{Bias}_J(\theta_i^{\widehat{W}})]^2,$$

$$\widehat{Var}_J(\theta_i^{\widehat{W}}) = \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} [\theta_{i(k)}^{\widehat{W}} - \theta_{i(\cdot)}^{\widehat{W}}]^2,$$

$$\log \theta_{i(k)}^{\widehat{W}} = \frac{\sum_{j=1}^{n_i} w_{ij} \log \widehat{\theta}_{ij} - w_{ik} \log \widehat{\theta}_{ik}}{w_i - w_{ik}},$$

$$\theta_{i(\cdot)}^{\widehat{W}} = \frac{1}{n_i} \sum_{k=1}^{n_i} \theta_{i(k)}^{\widehat{W}}$$

로 주어진다.

2.3 Mantel-Haenszel 추정량

Mantel과 Haenszel(1959)이 제안한 공통 오즈비의 추정량은 v_{ik} 를 가중치로 하여 다음과 주어진다.

$$\widehat{\theta}_i^{MH} = \frac{\sum_{k=1}^{n_i} x_{k1} (N_{k2} - x_{k2}) / N_k}{\sum_{k=1}^{n_i} x_{k2} (N_{k1} - x_{k1}) / N_k} = \frac{\sum_{k=1}^{n_i} v_{ik} \widehat{\theta}_{ik}}{\sum_{k=1}^{n_i} v_{ik}},$$

여기서 $v_{ik} = x_{k2} (N_{k1} - x_{k1}) / N_k$, $N_k = N_{k1} + N_{k2}$ 이다.

이 때 $\widehat{\theta}_i^{MH}$ 의 분산은 근사적으로 다음과 같이 주어진다.

$$\sigma^2(\widehat{\theta}_i^{MH}) = (\widehat{\theta}_i^{MH})^2 \frac{\sum_{k=1}^{n_i} v_{ik}^2 w_{ik}^{-1}}{v_i^2},$$

여기서 $v_i = \sum_{k=1}^{n_i} v_{ik}$ 이고 w_{ik} 는 2.2절에서 주어진 가중치와 같다.

따라서 소지역 i 에 대한 공통 오즈비의 Mantel-Haenszel 추정량($\widehat{\theta}_i^w$)의 편의와 평균제곱오차의 추정치는 근사적으로 다음과 같이 주어진다.

$$\widehat{Bias}(\widehat{\theta}_i^{MH}) = \widehat{\theta}_i^{MH} \left[\sum_{k=1}^{n_i} v_{ik} \exp\left(\frac{w_{ik}^{-1}}{2}\right) / v_i - 1 \right] \simeq \frac{\widehat{\theta}_i^{MH}}{2v_i} \sum_{k=1}^{n_i} v_{ik} w_{ik}^{-1}$$

$$\begin{aligned} \widehat{MSE}(\theta_i^{\widehat{MH}}) &= \widehat{Var}(\theta_i^{\widehat{MH}}) + [\widehat{Bias}(\theta_i^{\widehat{MH}})]^2 \\ &\simeq \frac{(\theta_i^{\widehat{MH}})^2}{v_i^2} \left[\sum_{k=1}^{n_i} v_{ik}^2 w_{ik}^{-1} + \frac{(\sum_{k=1}^{n_i} v_{ik} w_{ik}^{-1})^2}{4} \right]. \end{aligned}$$

또한, 앞에서와 마찬가지로 소지역 i 에서 공통 오즈비에 대한 Mantel-Haenszel 추정량의 편의와 MSE 추정치들은 잭나이프 추정방법을 이용하면 다음과 같이 주어진다.

$$\begin{aligned} \widehat{Bias}_J(\theta_i^{\widehat{MH}}) &= (n_i - 1)(\theta_{i(\cdot)}^{\widehat{MH}} - \theta_i^{\widehat{MH}}), \\ \widehat{MSE}_J(\theta_i^{\widehat{MH}}) &= \widehat{Var}_J(\theta_i^{\widehat{MH}}) + [\widehat{Bias}_J(\theta_i^{\widehat{MH}})]^2, \end{aligned}$$

여기서

$$\begin{aligned} \widehat{Var}_J(\theta_i^{\widehat{MH}}) &= \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} [\theta_{i(k)}^{\widehat{MH}} - \theta_{i(\cdot)}^{\widehat{MH}}]^2, \\ \theta_{i(k)}^{\widehat{MH}} &= \frac{\sum_{j=1}^{n_i} x_{j1}(N_{j2} - x_{j2})/N_j - x_{k1}(N_{k2} - x_{k2})/N_k}{v_i - v_{ik}} = \frac{\sum_{j=1}^{n_i} v_{ij}\hat{\theta}_{ij} - v_{ik}\hat{\theta}_{ik}}{v_i - v_{ik}}, \\ \theta_{i(\cdot)}^{\widehat{MH}} &= \frac{1}{n_i} \sum_{k=1}^{n_i} \theta_{i(k)}^{\widehat{MH}} \end{aligned}$$

로 주어진다.

3. 자료분석

대지역인 경기지역에 대해 조사된 2002년 12월 경제활동인구조사 자료를 이용하여 경기지역 내의 시군 단위 소지역들에 대한 추정결과를 설명하기로 한다. 경기지역은 시군단위의 31개 행정자치구역들로 구성되어 있다. 경제활동인구조사에서 경기지역에 할당된 표본조사구수는 118개이고, 조사가구수는 약 2,800개이다. 여기에서 표본조사구 수가 한 개만이 배정되어 있는 7개 소지역들은 분석에서 제외하였고, 나머지 24개 시군 단위 소지역들에 대해 남녀별 실업률의 공통 오즈비 추정을 실시하였다.

2002년 12월 경제활동인구조사 자료를 이용하여 추정된 경기지역 내의 24개 시군 단위 행정자치구역들에 대한 남녀별 실업률의 오즈비 추정결과가 다음 <표 1>에 주어졌다.

<표 1> 경기지역의 시군단위 소지역들에 대한 오즈비 및 편의와 MSE 추정치(2002년12월)

Area No.	합동추정량			Woolf 추정량			M-H 추정량		
	$\hat{\Theta}_i^p$	Bias	MSE	$\hat{\Theta}_i^w$	Bias	MSE	$\hat{\Theta}_i^{MH}$	Bias	MSE
1	1.14	0.14	0.26	1.01	0.05	0.10	1.05	0.07	0.13
2	1.13	0.29	0.59	1.00	0.15	0.27	1.09	0.17	0.34
3	1.62	-3.01	9.14	1.11	0.62	1.44	1.35	0.61	1.85
4	2.42	0.62	3.10	1.44	0.08	0.36	1.57	0.08	0.41
5	0.71	0.06	0.09	0.62	0.06	0.06	0.66	0.06	0.07
6	0.41	0.06	0.05	0.32	0.05	0.03	0.40	0.06	0.05
7	0.57	0.12	0.06	0.65	0.05	0.04	0.68	0.06	0.06
8	0.64	0.40	0.59	0.49	0.23	0.22	0.64	0.24	0.34
9	0.44	-0.08	0.08	0.47	0.53	1.04	0.47	0.58	1.12
10	0.76	0.10	0.19	0.71	0.06	0.09	0.69	0.09	0.12
11	5.10	-2.20	5.93	2.72	0.18	1.11	2.77	0.13	1.09
12	9.61	-3.98	27.25	4.66	0.96	12.30	5.17	0.45	11.58
13	1.33	0.34	0.72	1.08	0.10	0.17	1.14	0.08	0.20
14	1.09	0.26	0.40	1.09	0.17	0.30	1.21	0.19	0.43
15	0.46	0.09	0.06	0.52	0.10	0.09	0.53	0.08	0.09
16	0.72	0.75	2.09	0.69	0.25	0.49	0.68	0.34	0.63
17	1.73	0.10	0.09	1.74	-0.09	0.13	1.75	-0.08	0.12
18	1.15	0.09	0.24	0.99	0.05	0.11	1.06	0.06	0.14
19	0.68	0.02	0.04	0.55	0.13	0.10	0.57	0.12	0.10
20	0.98	-0.86	1.23	0.83	2.12	12.51	0.88	2.07	12.30
21	6.58	-4.81	32.35	1.52	0.06	0.43	1.71	0.01	0.46
22	1.13	-0.67	0.46	0.72	0.13	0.11	0.74	0.11	0.11
23	3.81	-1.46	4.73	2.09	0.26	2.67	2.27	0.08	2.60
24	5.93	-3.88	17.90	1.92	0.04	0.36	2.01	0.02	0.38

<표 1>을 살펴보면 남녀간 실업률의 오즈비에 대한 합동추정량의 MSE는 24개 행정자치구역 거의 대부분에서 다른 두 추정량들에 비해 다소 크게 나타나며 불안정하다. 반면, Woolf 추정량과 M-H 추정량의 MSE 추정치들은 합동추정량의 MSE에 비해 매우 작고 안정적이다. 또한 Woolf 추정량과 M-H 추정량의 편의 추정치들이 합동추정량에 비해 거의 모든 소지역들에서 다소 작게 나타나고 있다.

각 소지역들에 대한 합동추정량, Woolf 추정량, M-H 추정량들의 신뢰성을 평가하기 위해 상대편의(Relative Bias: RB)와 상대평균오차제곱근(Relative Root Mean Square Error: RRMSE)의 추정치를 계산하였다. 소지역 i 에 대한 남녀 실업률의 오즈비 추정량을 $\hat{\Theta}_i^*$ 로 나타낼 때, RB와 RRMSE는 각각 다음과 같이 주어진다.

$$RB(\hat{\theta}_i^*) = \frac{\widehat{Bias}(\hat{\theta}_i^*)}{\hat{\theta}_i^*} \times 100,$$

$$RRMSE(\hat{\theta}_i^*) = \frac{\sqrt{\widehat{MSE}(\hat{\theta}_i^*)}}{\hat{\theta}_i^*} \times 100.$$

경기지역의 경제활동인구조사 자료로부터 24개 행정자치구역들에 대한 남녀별 실업률의 오즈비 추정치들의 RB와 RRMSE 값들을 계산하면 다음 표 2와 같이 주어진다.

표 2에서 세 가지 추정치들의 편의 값들을 비교해 보면 Woolf 추정치들의 절대편의 값들의 평균(Av. RB)이 29.89%, M-H 추정치들의 경우에는 27.81%로 서로 비슷하게 나타나며 합동추정치들의 절대편의 값들의 평균인 41.02%보다 상대적으로 작게 나타난다. 그리고 소수를 제외한 거의 모든 행정자치지역들에 대해서 합동추정치들의 상대편의 값들은 비교적 큰 값을 나타낸다. 반면, Woolf 추정치들의 상대편의 값들은 Area 5, 6, 9, 19, 20을 제외한 나머지 19개 지역들에서 합동추정치들에 비해 매우 작고 안정적이고, M-H 추정치들의 경우에는 Area 9, 19, 20을 제외한 나머지 21개 지역들에서 합동추정치들에 비해 매우 작고 안정적이다. 또한 소지역들에 대한 상대편의(RB) 값들의 허용한계 기준을 40%로 할 때, 합동추정치들의 RB 값들은 14개 지역을 제외한 10개 지역에서 허용한계의 기준치를 만족하지 않지만, Woolf 추정치들의 경우에는 4개 지역(Area 3, 8, 9, 20)을 제외한 20개 소지역들에서 허용기준치를 만족하고, M-H 추정치들의 경우도 4개 지역(Area 3, 9, 16, 20)을 제외한 20개 소지역들에서 허용기준치를 만족한다. 그러나 10개 지역을 제외한 나머지 14개 지역들에 대해서는 Woolf 추정 및 M-H 추정치들의 상대편의 값들은 비교적 큰 값을 나타낸다. 따라서 이 추정량들의 신뢰성은 이들 추정량들의 편의와 분산의 영향이 함께 고려되어 평가되어야 한다.

남녀간 실업률의 오즈비에 대한 추정치들의 RRMSE 추정치를 통해 이들 추정량들의 신뢰성을 평가하고자 한다. Woolf 추정치들과 M-H 추정치들의 RRMSE 값들은 Area 9, 12, 15, 17, 19, 20, 23을 제외한 17개 소지역들에서 공통적으로 합동추정치들의 RRMSE 값들보다 작은 값을 나타내지만, 세 추정치들의 RRMSE 값들의 평균(Av. RRMSE)은 서로 비슷하게 나타난다. 따라서 대부분 소지역들에 있어서 Woolf 추정치나 M-H 추정치들이 합동추정치들에 비해 상대적으로 높은 안정성을 나타내며, 효율성 측면에서는 세 가지 추정치들의 결과가 거의 비슷하게 나타난다. 그러므로 현 경제활동인구조사 체계 하에서는 Woolf 추정량이나 M-H 추정량이 합동추정량에 비해 안정성은 높으나 효율성은 서로 비슷하다는 사실을 경기지역의 시군단위 소지역 추정 결과에서 알 수 있다.

<표 2> 경기지역의 시군단위 소지역들에 대한 오즈비의 RB와 RRMSE 추정치
(2002년 12월)

Area No.	합동추정량		Woolf 추정량		M-H 추정량	
	RB_i	$RRMSE_i$	RB_i	$RRMSE_i$	RB_i	$RRMSE_i$
1	12.63	44.97	5.18	31.11	6.43	34.01
2	25.21	68.04	14.83	52.17	15.69	53.77
3	-185.97	186.55	55.93	107.87	45.36	100.91
4	25.61	72.70	5.72	41.65	5.35	40.78
5	8.97	42.19	10.30	40.92	8.68	39.15
6	14.41	53.66	15.85	53.41	14.37	57.23
7	20.21	44.24	8.43	31.65	8.50	35.07
8	61.38	119.72	46.57	96.00	38.12	91.71
9	-18.73	64.64	112.82	216.08	124.62	226.79
10	12.69	56.51	8.18	41.76	12.59	50.82
11	-43.18	47.72	6.70	38.73	4.67	37.71
12	-41.45	54.31	20.67	75.22	8.78	65.78
13	25.27	63.51	9.02	37.92	6.83	39.18
14	23.76	58.44	16.00	50.75	15.33	54.37
15	19.39	53.80	18.83	58.18	15.36	55.81
16	105.04	201.79	36.31	101.43	50.30	116.92
17	5.92	17.84	-5.29	21.07	-4.73	20.16
18	8.17	47.28	4.79	32.61	5.24	34.88
19	2.25	27.78	24.23	57.92	21.68	55.13
20	-88.10	113.79	254.41	424.77	234.27	397.26
21	-73.08	86.43	4.07	43.28	0.62	39.83
22	-59.29	59.77	18.65	46.78	15.26	44.38
23	-38.28	57.02	12.42	77.98	3.61	71.04
24	-65.48	71.39	2.26	31.25	1.01	30.76
Av. RB	41.02		29.89		27.81	
Av. RRMSE		71.21		75.44		74.73

Av. RB = RB의 절대값들의 평균

Av. RRMSE = RRMSE 값들의 평균

4. 결론

우리나라의 경제활동인구조사는 대지역인 광역시 또는 도 단위들의 통계작성을 목적으로 실시되기 때문에 최근 들어 사회적 관심이 커지고 있는 시군구 단위의 행정자치구역들과 같은 소지역들에 대한 실업통계는 경제활동인구조사 자료만으로는 추정이 불가능하다. 그러나 이 연구에서는 경제활동인구조사 자료만을 이용하여 시군단위 행

정자치구역들에 대한 남녀별 실업률에 대한 오즈비를 추정하기 위해 합동추정량, Woolf 추정량과 Mantel-Haenszel 추정량을 고려하였다. 시군단위 소지역 추정치들의 안정성과 신뢰성의 측도로써 추정치들의 상대편의(RB)와 상대평균오차제곱근(RRMSE)이 제시되었다. 경제활동인구조사 자료를 이용한 추정결과에 따르면 시군단위 소지역 추정에 있어서 Woolf 추정량이나 Mantel-Haenszel 추정량이 합동추정량에 비해 보다 높은 안정성을 나타냈으며, 효율성의 측면에서는 세 추정량들이 서로 비슷하게 나타났다. 또한 이러한 추정량들은 대부분 소지역 추정치들의 상대편의 값들의 목표허용오차 한계를 만족하였다.

국내의 경제활동인구조사는 매월 주기적으로 실시되며 매 5년 단위로 표본개편이 이루어진다. 새로운 표본설계에서는 시군구단위 소지역 통계의 신뢰성이 확보될 수 있도록 모집단의 층화, 표본배정, 집락화의 수준 등에 대한 전반적인 검토가 요구된다. 또한 새로운 표본에서는 대지역 내의 시군구 단위 소지역들뿐만 아니라 성별, 연령별, 학력별 등과 같은 세부 관심영역들에 대한 시군구 단위 통계작성도 가능하도록 충분한 검토가 이루어져야 할 것이다.

참고문헌

1. Lee, Sang Eun(2001). Bayes Prediction for small area estimation, 한국통계학회논문집, 8-2, pp. 407-416.
2. Mantel, N. and Haenszel, W.(1959). Statistical aspects of the analysis of data from retrospective studies of disease, Journal of the National Cancer Institute, 22, pp. 719-748.
3. Woolf, B.(1955). On estimating the relation between blood group and disease, Annals of Human Genetics, 19, pp. 251-253.
4. 박종태, 이상은(2001). 소지역 추정법에 관한 비교연구, 국제데이터정보과학회지, 12-2, pp. 47-55.
5. 박종태, 광민정(2003). 경기지역의 실업률에 대한 오즈비의 비교연구, 평택대학교논문집, 17, pp. 305-313.
6. 양영춘, 이상은, 신민웅(2003). Kalman Filter 기법을 이용한 실업자 수의 소지역추정, 응용통계연구, 16-2, pp. 239-246.
7. 이상은, 진영(1996). 변동계수를 이용한 소지역 통계의 안정성 검토, 통계분석연구, 1, pp. 23-40.
8. 이계오(2000). 시군구 실업자 추정을 위한 소지역 추정법, 응용통계연구, 13-2, pp. 275-286.
9. 정연수, 이계오, 이우일(2003). 시군구 실업자 총계 추정을 위한 설계기반 간접추정법, 응용통계연구, 16-1, pp. 1-4.

[2006년 5월 접수, 2006년 6월 채택]