

## 소지역의 실업률에 대한 상대위험도의 추정에 관한 비교연구

박종태<sup>1</sup>

평택대학교 디지털응용정보학과

접수 2009년 1월 19일, 수정 2009년 3월 10일, 게재확정 2009년 3월 20일

### 요약

국내의 행정구역상 시군구 등과 같은 소지역에 있어서 실업률에 대한 남녀별 공통 상대위험도를 추정하는데, 추정방법으로 단순한 합동추정, 울프 방법에 기초한 가중추정과 잭나이프 추정들을 고려하고 이 추정 방법들의 효율성을 편의와 평균제곱오차의 개념을 통해서 비교하고자 한다. 이를 위해 2002년 12월 경기지역의 경제활동인구조사 자료를 이용하여 이 지역 내의 24개 시군단위 소지역들의 남녀별 실업률에 대한 상대위험도의 편의 및 평균제곱오차가 본 연구에서 제시된 추정절차에 의해 추정된다. 또한, 이들 추정치들의 안정성과 신뢰성은 상대편의와 상대오차제곱근을 통하여 비교된다. 추정결과 잭나이프 추정이 다른 두 추정들에 비해 매우 효율적임을 보였다.

주요용어: 상대오차제곱근, 상대위험도, 상대편의, 소지역, 평균제곱오차.

### 1. 서론

실업률은 국가경제와 밀접하게 관련된 변수중의 하나이다. 따라서 미국, 일본 등을 비롯하여 각 국가들은 세계적인 장기간의 경기 불황으로 인한 실업률을 낮추기 위해 인간 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 나아가 지역자치단체들도 예전과 달리 소지역통계, 특히 경제 관련통계를 철저히 필요로 하고 있다.

실업통계는 매월 정부주관의 노동력통계조사를 통하여 발표되고 있다. 그러나 실업률의 경우 우리나라에서는 현재 16개 시도단위의 대 지역에 대해서만 발표하고 있어 시군구단위의 소지역통계가 없다는 것이 안타까운 현실이다. 이를 극복하기 위한 국내의 작은 시도들을 소개하면 다음과 같다: Lee와 Jin (1996)은 1994년 경제활동인구조사 자료를 이용하여 충청도의 두 소지역에 있어서 실업률에 대한 3가지 추정량들을 변동계수로 비교하였고, Lee와 Park (1999)은 소지역에 대한 실업자 총수의 추정량을 모의실험을 통해 경험적으로 비교하였다. 또한 Lee (2000)는 인구통계학적 방법, 합성 추정, 복합 추정 등의 소지역 추정법을 이용하여 충북 시, 군, 구의 실업자 수를 추정하였다. 그리고 Lee (2001)는 소지역추정을 위해 베이스 예측값을 이용하였고, Park과 Lee (2001)는 직접추정, 합성추정, 복합추정, 베이스 추정 등의 효율성을 비교하고자 모의실험을 통해 편의와 평균제곱오차를 계산하였다. 또한 Yang 등 (2003)은 실업자 수의 소지역추정을 위해 Kalman Filter 기법을 이용하였고, Chung 등 (2003)은 시군

<sup>1</sup> (450-701) 경기도 평택시 용이동 111번지, 평택대학교 디지털응용정보학과, 부교수.  
E-mail: jtpark@ptu.ac.kr

구 실업자 총계 추정을 위해 표본설계에 기반을 둔 간접추정법을 제안하였다. 최근에 Park (2006)은 경기지역의 시군단위에 있어서 남녀별 실업률에 대한 오즈비의 추정을 위해 합동추정량, Woolf 추정량과 Mantel-Haenszel 추정량을 고려하고 이 추정량들의 안정성과 신뢰성을 편의와 평균제곱오차를 통해 비교하였다.

본 논문에서는 경기도의 각 시군을 소지역으로 하여 각 지역에 있어서 남녀별 실업률의 상대위험도(relative risk)를 그 소지역의 전체자료(pooled data)를 이용한 합동추정량(pooled estimator), Woolf (1955) 방법에 기초한 가중추정량, 잭나이프 방법 (Miller, 1974)에 의한 추정량 등의 방법으로 추정하고 이들의 신뢰성을 편의(Bias)와 평균제곱오차(MSE)를 통해 비교하고자 한다. 2002년 12월 경기지역의 경제활동인구조사 자료로부터 24개 시군단위 소지역들에 대한 남녀 각각의 표본 실업률과 남녀 그룹별 상대위험도를 추정한다. 또한 시군단위 소지역에 있어서 상대위험도의 추정치들의 신뢰성은 이들 추정치들의 상대편의(RB: relative bias)와 상대오차제곱근(RRMSE: relative root mean square error)에 의해 비교된다.

### 2. 소지역의 실업률에 대한 공통 상대위험도의 추정

국내의 행정구역상 도나 광역시에 해당하는 어떤 대지역이  $I$ 개의 시군구 등과 같은 소지역으로 분할되었다고 가정하자. 이 때 소지역  $i$ 에 할당된 표본조사구의 총수를  $N_{ik}$  라고 하고 이 중에서  $k$ 번째 표본 조사구에서 임의로 추출된 성별변수와 경제활동변수의 값에 따라 다음 표 2.1과 같이  $2 \times 2$ 분할표로 분류되었다고 가정하자.

표 2.1 성별과 경제활동상태에 따른 분할표

성별	실업	취업	합계
1(남성)	$x_{ik1}$	$N_{ik1} - x_{ik1}$	$N_{ik1}$
2(여성)	$x_{ik2}$	$N_{ik2} - x_{ik2}$	$N_{ik2}$

여기서  $N_{ik1}$ (또는  $N_{ik2}$ )은  $i$ 번째 소지역 내의  $k$ 번째 조사구에서 경제활동을 하고 있는 남성 (또는 여성)의 인구수이고,  $x_{ik1}$ (또는  $x_{ik2}$ )은 그 조사구에서 실업자로 조사된 남성 (또는 여성)의 인구수이다.

이 때 소지역  $i$ 의  $k$ 번째 표본 조사구에 대한 실업률의 상대위험도(relative risk)는 두 그룹의 실업률의 비로서 다음과 같이 정의된다.

$$r_{ik} = \frac{\pi_{ik1}}{\pi_{ik2}}$$

여기서  $\pi_{ik1}$ 은  $i$ 번째 소지역의  $k$ 번째 표본조사구의 남성에 대한 실업률이고  $\pi_{ik2}$ 는 여성에 대한 실업률을 나타낸다.

또한  $p_{ik1}$ 과  $p_{ik2}$ 를 표본실업률로 갖는 두 그룹의 표본상대위험도는

$$\hat{r}_{ik} = \frac{p_{ik1}}{p_{ik2}} = \frac{x_{ik1}/N_{ik1}}{x_{ik2}/N_{ik2}}$$

로 정의되며, Agresti (1996)에 의하면 로그상대위험도인  $\log \hat{r}_{ik}$ 의 표본분포는 만일 표본크기가 커지면 평균이  $\log r_{ik}$ 이고 다음과 같은 근사적 표준편차를 갖는 정규분포로 수렴한다고 알려져 있다.

$$\hat{\sigma}(\log \hat{r}_{ik}) \simeq \sqrt{\frac{1 - p_{ik1}}{N_{ik1}p_{ik1}} + \frac{1 - p_{ik2}}{N_{ik2}p_{ik2}}}$$

이제 주어진 소지역의 모든 표본 조사구에서 남녀별 실업에 대한 상대위험도가 모두 동일하다고 가정한다. 즉,  $r_{i1} = r_{i2} = \dots = r_{in_i} = r_i$ 로 가정하자.

이 때 소지역의 공통 상대위험도인  $r_i$ 를 추정하기 위해 그 소지역의 조사구 전체에 대한 단순 합동추정, Woolf 방법에 기초한 가중추정 그리고 잭나이프(Jackknife) 추정 방법을 고려하기로 한다.

**2.1. 합동추정**

소지역  $i$ 에서의 공통 상대위험도의 단순한 합동추정량(pooled estimator)은 그 소지역의 모든 조사구를 합한 전체의 남성과 여성의 표본 실업률의 비로 다음과 같이 간단하게 나타낼 수 있다.

$$\hat{r}_i^p = \frac{x_{i1}/N_{i1}}{x_{i2}/N_{i2}}.$$

여기서  $x_{i1} = \sum_{k=1}^{n_i} x_{k1}$ ,  $x_{i2} = \sum_{k=1}^{n_i} x_{k2}$ ,  $N_{i1} = \sum_{k=1}^{n_i} N_{k1}$ ,  $N_{i2} = \sum_{k=1}^{n_i} N_{k2}$ 이다.

마찬가지로 대표본일 경우, 이 상대위험도의 로그 변환  $\log \hat{r}_i^p$ 의 표준편차는 다음과 같이 근사적으로 주어진다 (Agresti, 1996).

$$\hat{\sigma}(\log \hat{r}_i^p) \simeq \sqrt{\frac{1-p_{i1}}{N_{i1}p_{i1}} + \frac{1-p_{i2}}{N_{i2}p_{i2}}}.$$

여기서,  $p_{i1} = x_{i1}/N_{i1}$ ,  $p_{i2} = x_{i2}/N_{i2}$ 이다.

뿐만 아니라, 로그 상대위험도의 표본분포는 표본크기가 커지면 정규분포에 근사하므로 상대위험도의 표본분포는 근사적으로 로그정규분포(log normal)에 수렴하게 된다. 따라서 소지역  $i$ 에 대한 공통 상대위험도의 합동추정량의 편의와 평균제곱오차의 추정량은 간단한 수학적 계산에 의해 다음과 같이 근사적으로 나타낼 수 있다 (Park, 2006).

$$\begin{aligned} \text{Bias}(\hat{r}_i^p) &= \hat{r}_i^p \left[ \exp\left(\frac{1}{2}\hat{\sigma}^2(\log \hat{r}_i^p)\right) - 1 \right] \simeq \frac{1}{2}\hat{r}_i^p \left[ \frac{1-p_{i1}}{N_{i1}p_{i1}} + \frac{1-p_{i2}}{N_{i2}p_{i2}} \right] \\ \text{MSE}(\hat{r}_i^p) &= (\hat{r}_i^p)^2 \left[ \exp(2\hat{\sigma}^2(\log \hat{r}_i^p)) - 2\exp\left(\frac{1}{2}\hat{\sigma}^2(\log \hat{r}_i^p)\right) + 1 \right] \\ &\simeq (\hat{r}_i^p)^2 \left[ \left(\frac{1-p_{i1}}{N_{i1}p_{i1}} + \frac{1-p_{i2}}{N_{i2}p_{i2}}\right) + \frac{7}{4}\left(\frac{1-p_{i1}}{N_{i1}p_{i1}} + \frac{1-p_{i2}}{N_{i2}p_{i2}}\right)^2 \right]. \end{aligned}$$

**2.2. 가중추정**

소지역  $i$ 에서의 공통 상대위험도의 로그 가중추정량은 Woolf 방법에 기초하여 근사적 분산의 역수를 가중치로 각각의 로그 상대위험도의 합으로 다음과 같이 주어진다.

$$\log \hat{r}_i^w = \frac{\sum_{k=1}^{n_i} w_{ik} \log \hat{r}_{ik}}{\sum_{k=1}^{n_i} w_{ik}}.$$

여기서  $w_{ik}^{-1} = \hat{\sigma}^2(\log \hat{r}_{ik}) = (1-p_{ik1})/N_{ik1}p_{ik1} + (1-p_{ik2})/N_{ik2}p_{ik2}$ 이다. Woolf (1955)에 의하면 대표본일 경우,  $\log \hat{r}_i^w$ 의 분산은 다음과 같이 근사적으로 주어진다.

$$\hat{\sigma}^2(\log \hat{r}_i^w) \simeq \frac{1}{w_i},$$

여기서  $w_i = \sum_{k=1}^{n_i} w_{ik}$ 이다. 따라서 합동추정량의 경우와 마찬가지로 소지역  $i$ 에 대한 공통 상대위험도의 가중추정량의 편의와 평균제곱오차는 간단한 수학적 계산에 의해 다음과 같이 근사적으로 나타낼 수 있다 (Park, 2006).

$$\begin{aligned}\hat{\text{Bias}}(r_i^w) &= r_i^w \left[ \exp\left(\frac{1}{2}\hat{\sigma}^2(\log r_i^w)\right) - 1 \right] \simeq r_i^w \frac{w}{2w_i}, \\ \hat{\text{MSE}}(r_i^w) &= (r_i^w)^2 \left[ \exp(2\hat{\sigma}^2(\log r_i^w)) - 2\exp\left(\frac{1}{2}\hat{\sigma}^2(\log r_i^w)\right) + 1 \right] \\ &\simeq (r_i^w)^2 \left[ \frac{1}{w_i} + \frac{7}{4w_i^2} \right].\end{aligned}$$

### 2.3. 잭나이프 추정

여기서는 2.2절에서 언급한 가중추정량에 대하여 공통 상대위험도의 잭나이프 방법을 고려하자. 소지역  $i$ 에 있어서 안정된 분산 추정치를 얻기 위해 가중 추정량에 로그를 취하고 표본 조사구를 층으로 하여 층을 하나씩 제거하는 방법으로 유사-값(pseudo-value)  $r_{ik}^w$ 을 계산한다. 즉,

$$r_{ik}^w = n_i \log r_i^w - (n_i - 1) \log r_{i(k)}^w.$$

여기서,  $\log r_{i(k)}^w = (\sum_{j=1}^{n_i} w_{ij} \log r_{ij}^w - w_{ik} \log r_{ik}^w) / (w_i - w_{ik})$ 이다. 이 때 공통 상대위험도의 로그 잭나이프 추정량은 다음과 같이 주어진다 (Miller, 1974).

$$\log \hat{r}_i^J = \frac{1}{n_i} \sum_{k=1}^{n_i} r_{ik}^w.$$

또한, 잭나이프 추정량의 편의와 분산, 그리고 평균제곱오차는 각각 다음과 같이 근사적으로 나타낼 수 있다.

$$\begin{aligned}\hat{\text{MSE}}(r_i^J) &\simeq [\hat{\text{Bias}}(r_i^J)]^2 + \hat{\text{Var}}(r_i^J), \\ \hat{\text{Bias}}(r_i^J) &\simeq (n_i - 1)(\hat{r}_i(\cdot)^w - r_i^w) = \hat{r}_i^w - r_i^J, \\ \hat{\text{Var}}(r_i^J) &\simeq (r_i^J)^2 \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} [r_{i(k)}^w - r_{i(\cdot)}^w]^2.\end{aligned}$$

여기서,  $r_i(\cdot)^w = \sum_{k=1}^{n_i} r_{i(k)}^w / n_i$ 이다.

## 3. 자료 분석 결과

도 단위인 경기도 지역에 대해 조사된 2002년 12월 경제활동인구조사 자료를 이용하여 경기지역 내의 시군단위 소지역에 있어서 실업률에 대한 상대위험도의 세 가지 추정방법을 비교하고자 한다. 경기지역은 시군단위의 31개 행정자치구역들로 구성되어 있다. 경제활동인구조사에서 경기지역에 할당된 표본조사구수는 118개이고, 조사가구 수는 약 2,800개이다. 여기에서 표본조사구 수가 한 개만이 배정되어 있는 7개 소지역들은 분석에서 제외하였고, 나머지 24개 시군단위 소지역들에 대해 남녀별 실업

률의 공통 상대위험도 추정을 실시하였다. 2002년 12월 경제활동인구조사 자료를 이용하여 추정된 경기지역 내의 24개 시군단위 소지역들에 대한 남녀별 실업률에 대한 상대위험도의 추정결과는 다음 표 3.1에 주어져 있다.

표 3.1을 살펴보면 남녀 간 실업률의 상대위험도에 대한 합동추정치들과 가중추정치들의 MSE는 24개 소지역 거의 대부분에서 크게 나타나며 불안정하다. 반면, 잭나이프 추정치들의 MSE 값들은 합동추정치나 가중추정치들의 MSE에 비해 매우 작고 안정적이다. 세 가지 추정량들 중 잭나이프 추정치들의 MSE 값들이 가장 작고 안정적으로 나타나고 있다.

또한, 각 소지역들에 대한 합동추정량, 가중추정량, 잭나이프 추정량들의 신뢰성을 평가하기 위해 상대편의(RB)와 상대오차제곱근(RRMSE)의 추정치를 계산하였다.  $i$ 번째 소지역에 있어서 남녀 실업률에 대한 상대위험도의 일반적인 추정량을  $\hat{r}_i$ 로 나타낼 때, RB와 RRMSE는 각각 다음과 같이 주어진다.

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

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

표 3.1 경기지역의 시군단위 소지역들에 대한 상대위험도 및 MSE

Area No.	합동추정		가중추정		잭나이프추정	
	$\hat{r}_i^p$	MSE	$\hat{r}_i^w$	MSE	$\hat{r}_i^j$	MSE
1	1.13	0.23	1.00	0.18	1.00	0.07
2	1.12	0.25	0.97	0.20	0.92	0.14
3	1.56	1.82	1.05	0.61	0.79	0.34
4	2.33	3.80	1.40	0.80	1.44	0.30
5	0.75	0.06	0.65	0.06	0.62	0.04
6	0.46	0.03	0.38	0.02	0.35	0.02
7	0.59	0.25	0.68	0.26	0.67	0.03
8	0.67	0.15	0.50	0.10	0.43	0.07
9	0.46	0.22	0.50	0.25	0.38	0.17
10	0.78	0.14	0.74	0.14	0.74	0.06
11	4.75	234.90	2.55	22.23	2.54	0.82
12	8.04	602.71	3.90	46.50	4.05	5.20
13	1.31	2.76	1.07	1.66	1.05	0.14
14	1.07	0.19	1.03	0.18	0.99	0.06
15	0.51	0.05	0.60	0.06	0.56	0.06
16	0.74	0.21	0.72	0.21	0.69	0.16
17	1.56	0.36	1.51	0.28	1.60	0.05
18	1.13	0.37	0.99	0.27	1.01	0.08
19	0.69	0.63	0.57	0.33	0.50	0.05
20	0.98	1.74	0.85	2.37	0.92	2.70
21	6.32	412.70	1.49	2.72	1.58	0.45
22	1.12	4.40	0.73	0.95	0.64	0.05
23	3.59	152.23	1.99	19.72	2.44	3.76
24	5.59	328.83	1.84	4.88	1.89	0.29

경기지역의 경제활동인구조사 자료로부터 24개 소지역들에 대한 남녀별 실업률에 대한 상대위험도의 추정치들의 RB와 RRMSE 값들을 계산하면 다음 표 3.2와 같이 주어진다.

표 3.2에서 세 추정치들의 편차 값을 비교해 보면 잭나이프 추정치의 RB 값들의 평균(Av. RB)이 8.12%로 합동추정치나 가중추정치보다 현저히 작은 값을 나타낸다. 그리고 소수를 제외한 거의 모든 소지역들에 대해서 합동추정 및 가중추정치의 상대편의 값들은 비교적 큰 값을 나타낸다. 반면, 잭나이프 추정치의 상대편의 값들은 Area 3, 8, 9, 15를 제외한 나머지 20개 소지역들에서 두 추정치들에 비해 매우 작고 안정적임을 알 수 있다.

표 3.2 경기지역의 시군단위 소지역들에 대한 상대위험도의 RB와 RRMSE

Area No.	합동추정		가중추정		잭나이프추정	
	RB <sub>i</sub>	RRMSE <sub>i</sub>	RB <sub>i</sub>	RRMSE <sub>i</sub>	RB <sub>i</sub>	RRMSE <sub>i</sub>
1	7.24	42.59	7.15	42.27	0.64	25.78
2	7.87	44.79	8.16	45.81	5.19	40.17
3	21.36	86.41	17.19	74.22	32.61	74.15
4	20.44	83.74	13.82	64.03	-2.64	37.87
5	4.59	32.66	5.59	36.56	3.97	32.05
6	5.38	35.77	6.01	38.14	5.92	36.53
7	20.82	84.85	17.43	74.92	2.25	24.38
8	11.97	58.30	13.20	62.12	18.52	60.87
9	27.59	104.15	26.01	99.70	30.16	107.26
10	8.70	47.64	9.65	50.81	0.00	32.54
11	108.38	322.34	57.16	185.21	0.35	35.80
12	102.03	305.41	53.32	174.82	-3.59	56.36
13	35.62	126.51	33.41	120.39	1.91	35.03
14	6.61	40.36	6.99	41.73	3.43	24.57
15	7.79	44.52	6.93	41.50	8.48	45.24
16	12.99	61.49	13.81	64.02	4.41	57.44
17	6.13	38.61	5.28	35.36	-5.88	13.64
18	10.41	53.29	10.16	52.47	-1.25	27.69
19	31.68	115.58	26.12	99.99	15.33	43.72
20	38.72	135.02	55.34	180.28	-7.61	177.93
21	107.95	321.19	29.94	110.74	-5.69	42.51
22	57.93	187.29	38.16	133.52	13.83	35.50
23	116.26	343.33	71.19	222.96	-18.31	79.58
24	109.13	324.34	33.29	120.06	-2.77	28.59
Av. RB	36.98		23.55		8.12	
Av. RRMSE		126.67		90.48		48.97

남녀별 실업률에 대한 상대위험도의 추정치들의 RRMSE 값을 통해 이들 추정량들의 신뢰성을 평가하고자 한다. 잭나이프 추정치의 RRMSE 값들은 Area 6, 8, 9, 15, 20을 제외한 나머지 19개 소지역들에서 합동추정이나 가중추정치의 RRMSE 값들보다 현저히 작은 값을 나타내며, 전체적으로 세 추정치들의 RRMSE 값들의 평균(Av. RRMSE)을 비교해도 잭나이프 추정치의 경우가 48.97%로 상대적으로 가장 작게 나타났음을 알 수 있다. 이 결과로 미루어 볼 때 거의 대부분 소지역들에 있어서 잭나이프 추정이 상당히 신뢰할 만한 추정 결과를 나타내며, 효율성 측면에서도 합동추정이나 가중추정에 비해 상대적으로 높은 효율성을 나타낸다. 따라서 잭나이프 추정이 다른 두 추정들에 비해 안정성과 신뢰성이 매우 높다는 사실을 경기지역의 시군단위 소지역 추정결과에서 알 수 있다.

#### 4. 결론 및 토론

이 연구에서는 시군단위 소지역들에 있어서 남녀별 실업률에 대한 상대위험도의 추정방법을 비교하기 위해 단순한 합동추정, 울프 방법에 기초한 가중추정과 잭나이프 추정 방법을 고려하였다. 그리고 이 추

정치들을 비교하기 위하여 평균제곱오차와 상대오차제곱근이 고려되었다. 경제활동인구조사 자료를 이용한 추정결과에 따르면 시군단위 소지역 추정을 비교해 볼 때 거의 대부분 소지역들에 있어서 잭나이프 추정이 합동추정이나 가중추정에 비해 월등한 효율을 나타냈으며, 전체적으로 잭나이프 추정이 다른 두 추정들에 비해 탁월하였다.

국내의 경제활동인구조사는 6대 광역시 또는 도 단위의 광역자치단체에 대한 통계작성을 목적으로 실시되기 때문에 최근 들어 사회적 관심이 커지고 있는 시군구 단위의 기초자치단체와 같은 소지역들에 대한 실업통계는 현재의 경제활동인구조사 시스템 하에서는 추정이 상당히 어렵고 거의 불가능하다 해도 과언이 아닐 것이다. 또한 현재 발표되고 있는 시도단위의 실업률은 실업자 분류기준을 감안하더라도 학계내의 관련 전문가뿐만 아니라 일반인들이 느끼는 체감지수와도 상당한 괴리가 있음이 늘 지적되고 있다. 이러한 현상은 예를 들어, 국내 금융위기 이후 고용환경의 변화로 장기 실업상태로 말미암은 구직 단념으로 인해 사실상 실업 인구가 비경제활동인구로 분류되는 등 여러 가지 요인이 있을 것이라 생각된다. 현재 나타나고 있는 결점 등을 시급히 보완하고 개선하기 위해서라도 다양한 분류기준에 따른 실업 통계 작성이 가능하도록 이에 대한 지속적인 연구가 선행되어야 할 것이다.

### 참고문헌

- Agresti, A. (1996). *An introduction to categorical data analysis*, Wiley, New York.
- Chung, Y. S., Lee, K. O. and Lee, W. I. (2003). Design-based small area estimation for the Korean economically active population survey. *The Korean Journal of Applied Statistics*, **16**, 1-4.
- Lee, K. O. (2000). On application of small area estimation to the unemployment statistics of si-gun-gu. *The Korean Journal of Applied Statistics*, **13**, 275-286.
- Lee, S. E. (2001). Bayes prediction for small area estimation. *The Korean Communications in Statistics*, **8**, 407-416.
- Lee, S. E. and Jin, Y. (1996). Evaluating stability of small area estimates with CVs in 1994 EAP survey. *Journal of Statistical Analysis*, **1**, 23-40.
- Lee, S. E. and Park, J. T. (1999). Empirical study on estimating the totals of small area. *Bulletin of the International Statistical Institute in 52nd ISI Conference*, 213-214.
- Miller, R. G. (1974). The Jackknife - a review. *Biometrika*, **61**, 1-15.
- Park, J. T. (2006). On estimating the odds ratio between male and female unemployment rate in small area. *Journal of the Korean Data & Information Science Society*, **17**, 1029-1039.
- Park, J. T. and Lee, S. E. (2001). A comparative study of small area estimation methods. *Journal of the Korean Data & Information Science Society*, **12**, 47-55.
- Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics*, **19**, 251-253.
- Yang, Y. C., Lee, S. E. and Shin, M. W. (2003). Small area estimation of unemployment using Kalman Filter method. *The Korean Journal of Applied Statistics*, **16**, 239-246.

## A comparison study on the estimation of the relative risk for the unemployed rate in small area

Jong Tae Park<sup>1</sup>

Department of Digital Information and Statistics, Pyeongtaek University

Received 19 January 2009, revised 10 March 2009, accepted 20 March 2009

### Abstract

In this study, we suggest the estimation method of the relative risk for the unemployment statistics of a small area such as si, gun, gu in Korea. The considered method are the usual pooled estimator, weighted estimator with the inverse of log-variance as weights, and the Jackknife estimator. And we compare with the efficiency of the three estimators by estimating the bias and mean square errors using real data from the 2002 Economically Active Population Survey of Gyeonggi-do. We compute the unemployed rate of male and female in small areas, and then estimate the common relative risk for the unemployed rate between male and female. Also, the stability and reliability of the three estimators for the common relative risk was evaluated using the RB(relative bias) and the RRMSE(relative root mean square error) of these estimators. Finally, the Jackknife estimator turned out to be much more efficient than the other estimators.

*Keywords:* Mean square error, relative bias, relative risk, relative root mean square, small area.

---

<sup>1</sup> Associate Professor, Department of Digital Information and Statistics, Pyeongtaek University, Pyeongtaek 450-701, Korea. E-mail: jtpark@ptu.ac.kr