

최종 보고서

소지역 추정법에 의한 시·군·구 실업통계 개발

2002년 12월

이 계 오 교수

목 차

제1장 서론	1
제2장 경제활동 통계 작성 현황	3
2.1 표본설계 및 추정	3
2.2 표본관리 및 오차관리	7
2.3 소지역(시군구)추정 필요성	9
제3장 소지역 추정법에 의한 노동통계 작성사례	11
3.1 미국	11
3.1.1 서론	11
3.1.2 불편성조정	13
3.1.3 무응답에 대한 보정	16
3.1.4 비 보정	17
3.1.5 복합가중치	22
3.1.6 계절요인 보정	25
3.1.7 주(state) 및 주 내의 부차관심영역에 대한 추정	26
3.1.8 적용 사례	32
3.2 캐나다	57
3.2.1 서론	57
3.2.2 층화 및 추출단위 구성	62
3.2.3 표본배정, 추출, 순환	66
3.2.4 특별조사와 보충조사	68
3.2.5 가중치와 추정	70
3.2.6 데이터 관리	81
3.2.7 소지역 추정법	88
3.3 영국	138
3.3.1 서론	138
3.3.2 소지역 추정방법	139
3.4 프랑스	142
3.5 일본	144
3.5.1 서론	144

3.5.2	층화 및 추출단위 구성	146
3.5.3	표본배분, 추출, 교체방식	151
3.5.4	추정방법	158
3.5.5	오차 관리	163
제4장	소지역 추정법	167
4.1	개 요	167
4.2	인구통계학적 방법	168
4.2.1	생명률법	168
4.2.2	성분법	169
4.2.3	회귀정후법	169
4.3	합성추정법	170
4.4	복합추정법	172
4.5	모형기반 간접추정법	174
4.5.1	기본적인 지역수준모형	174
4.5.2	경험적 배이즈방법	175
4.5.3	계층적 배이즈방법	180
제5장	시군구 경찰통계 작성	183
5.1	개 요	183
5.2	소지역추정을 위한 집락화 방안	184
5.3	직접추정법	205
5.4	합성추정법	206
5.5	복합추정법	211
5.6	SAS 프로그램 알고리즘	214
5.6.1	프로그램 순서도	214
5.6.2	세부 알고리즘	215
5.7	추정결과	232
5.7.1	시군구 추정결과	232
5.7.2	효율비교	246
제6장	시·도 단위 세부영역 통계 작성	259
6.1	개 요	259
6.2	특·광역시 및 도지역 집락화 방안	260
6.3	소지역추정법 적용	263
6.4	SAS 프로그램 알고리즘	269
6.5	성별-연령대별 범주 추정결과	281

6.6 성별-교육정도별 범주 추정결과	291
제7장 결 언	301
참고문헌	304
부록1 시군구 추정결과(전국)	307
(실업자 총계, 취업자 총계, 경제활동인구, 비경제활동인구, 실업률 추정결과 포함)	
부록2 시·도단위 세부영역 추정결과(전국)	320
(실업자 총계, 취업자 총계, 경제활동인구, 비경제활동인구, 실업률 추정결과 포함)	
부록3 시군구 추정 SAS프로그램	337
부록4 시·도단위 세부영역 추정 SAS 프로그램	356

제1장 서론

정보통신기술의 획기적인 발달로 사회는 글로벌 정보화로 발전하면서 다양한 분야에서 통계정보의 활용이 증대되고 있을 뿐 아니라 통계에 대한 신뢰성, 정확성과 시의성 등 다양한 특성들이 요구되고 있으며 더욱이 1995년에 시작된 지방자치제도의 활성화로 시군구 단위에 대한 지역통계의 필요성이 대두되고 있다. 특히 해당지역의 경제활동인구 관련 통계의 요구는 빈번할 뿐 아니라 지방자치 단체장들의 주요 관심 통계로 부각되어가고 있다. 그러나 현재와 같은 통계청의 인력과 조직으로 시군구 단위까지 경제활동인구 관련 통계를 생산하는 것은 어렵기 때문에 현재 주어진 여건을 감안하여 경제적으로 큰 부담 없이 일정 수준의 정확성을 갖춘 시군구 단위의 경제활동인구 관련 통계를 생산할 수 있는 방안에 대한 연구가 요구되고 있다.

또한 실업 관련 통계가 사회적으로 주요 관심사가 되면서 특·광역시와 도 단위의 세분화된 통계 소요도 제기되고 있다. 예를 들면 현재 통계청에서는 특·광역시와 도 단위에 대한 노동력 통계의 기본적인 사항인 취업자 총계, 실업자 총계와 비경제활동인구 등을 생산하고는 있지만 좀더 세분화된 항목인 연령계층별, 교육정도별 등에 대한 통계를 생산하지 않음으로써 광역자치단체별로 지역적 특성에 부합하고 효율적인 고용증진 정책을 수립하고 추진하는데 필수적인 연령대별과 교육정도별 경제활동인구 관련 통계의 생산에 대한 요구를 충족시키지 못하고 있다. 따라서 현재 우리나라의 통계 작성 체계 하에서 표본가구 수를 획기적으로 증대시키지 않으면서 경제적으로 특·광역시와 도 단위의 세분화된 항목에 대한 통계를 생산할 수 있는 소지역 추정법에 대한 연구가 절실하게 필요하다.

특히 금년에는 2000년 「인구주택 총 조사」 자료를 토대로 가구조사 표본을 전면 개편하여 병행조사를 실시하였다. 이러한 신규표본을 근거로 소지역 추정기법을 적용하여 시군구 단위의 경제활동인구 관련 통계와 특·광역시와 도 단위의 연령계층별과 교육정도별 취업자 총계, 실업자 총계와 비경제활동인구 통계를 생산할 수 있는 기틀을 마련한다면 2003년부터 신규표본에 의한 시군구 단위 소지역 통계 및 특·광역시와 도 단위 세부영역 통계의

생산이 가능할 것이다. 만일 금년에 이와 같은 연구를 시도하지 않는다면 앞으로 또 다른 5년이 지나서야 시군구 단위에 대한 실업통계의 생산이 가능할 것으로 사료되므로 소지역 추정법을 적용한 시군구 단위의 경제활동인구 통계와 특·광역시와 도 단위의 세부영역 통계작성 방안의 연구는 매우 중요한 과제라고 생각된다.

본 연구보고서는 소지역 추정법을 이용하여 시군구 단위 및 특·광역시와 도 단위의 연령대별과 교육정도별 경제활동인구 통계 작성 기법에 대한 내용뿐만 아니라 실제 표본조사 자료로부터 경제활동인구 관련 통계를 생산할 수 있는 알고리즘 개발에 대한 내용을 포함한다. 연구 결과를 실무에 적용하고 새로운 연구의 기본으로 활용할 수 있도록 새로 개편되는 경제활동인구 통계 작성 체계를 정리 요약하였고, 미국, 캐나다, 영국, 프랑스와 일본 등의 통계 선진국에서 경제활동인구 관련 통계분야의 소지역 추정법 활용실태를 정리 분석하여 본 연구방향 설정과 알고리즘을 연구 개발하는데 참고하였으며, 소지역 추정법의 일반적인 이론과 시군구 단위 경제활동인구 통계 작성에 필요한 절차를 연구 정리하였다. 또한 2003년에 적용되는 새로운 표본설계 하에서 시군구 단위의 경제활동인구 관련 통계를 작성할 수 있는 알고리즘을 연구하였으며, 알고리즘은 SAS언어로 구현하고 실제 계산결과를 통해서 타당성과 실용성을 검토하였다. 마지막으로 특·광역시 및 도 단위의 연령대별과 교육정도별 세부항목의 통계를 작성하기 위한 추정법의 적용절차와 계산 알고리즘을 개발하여 이를 SAS로 구현하고 수치적인 비교를 통해 타당성과 신뢰성을 입증하였다.

제2장 경제활동 통계작성 현황

2.1 표본설계 및 추정

경제활동인구조사의 표본설계는 매 5년마다 개편된다. 제 2장에서 언급할 표본설계는 2000년에 실시한 「인구주택 총 조사」를 근거로 작성된 내용이며 2003년부터 전면 개편할 내용이다. 신규 표본설계는 2000년 「인구주택 총 조사」에서 10% 표본조사구를 표본추출틀로 사용하였으며 실제로 보통조사구와 아파트조사구만을 정리한 결과 24,998조사구가 표본추출틀이 되었다.

조사모집단은 25개 층(7개 특광역시와 9개 도의 동부와 읍면부)으로 구분하고, 각 층의 특성을 분석하여 25개 층을 2개 그룹으로 다음과 같이 분류하였다.

그룹 1 (11개 층)

광주 광역시, 대전 광역시, 경기도 동부, 경기도 읍면부, 강원도 읍면부, 충남 읍면부, 전남 읍면부, 경남 동부, 경남 읍면부, 제주도 동부, 제주도 읍면부

그룹 2 (14개 층)

서울 특별시, 부산 광역시, 대구 광역시, 인천 광역시, 울산 광역시, 강원도 동부, 충북 동부, 충북 읍면부, 충남 동부, 전북 동부, 전북 읍면부, 전남 동부, 경북 동부, 경북 읍면부

전국의 표본조사구 수는 조사원의 인적능력을 감안하여 1,629개로 정하였고, 이들을 25개 층에 적절히 배분하고 각 층 내에서 대표성 있는 조사구를 선정하기 위해 두개 그룹에 대해 분류지표를 다르게 적용하여 조사구들을 정렬한 후, 계통추출법으로 표본조사구를 추출하였다. 두개 그룹

에 대한 분류지표는 다음과 같다.

그룹 1의 분류지표 (18가지)

- 주택 특성 : 아파트 1, 단독 주택 2, 다가구연립 기타 3
- 산업 특성 : 농림어업 1, 광공업 2, 서비스업 3
- 승용차와 컴퓨터 보유비율 : 평균보다 많음 1, 작음 2

그룹 2의 분류지표 (36가지)

- 주택 특성 : 아파트 1, 단독 주택 2, 다가구연립 기타 3
- 산업 특성 : 농림어업 1, 광공업 2, 서비스업 3
- 승용차와 컴퓨터 보유비율 : 평균보다 많음 1, 작음 2
- 실업자 수(15세이상 인구 중 실업자 비율의 층별 평균) :
 큼 1, 적음 2

추출된 표본과 모집단의 특성에 대한 유사성을 검토하기 위해 100회 정도의 표본추출 모의실험과정을 통해 모집단의 특성에 가장 근접하는 표본을 선정하였다. 이번에 개편된 표본설계는 일종의 연동표본가구 관리체계가 적용되었다. 표본가구로 선정된 가구는 첫 일년 동안은 교체 없이 지속적으로 조사를 실시하고 일년이 지나면 $\frac{1}{36}$ 씩 매월 새로운 표본으로 교체하는 형식이다.

본 절에 요약된 추정식은 첫 일년동안 고정표본가구 관리체계에서 적용될 것이며 다음과 같다. 경제활동의 특성 중에서 경제활동인구와 실업자 총계만을 추정한다면 실업률은 계산될 수 있으므로 경제활동인구, 실업자 총계 또는 비경제활동인구를 변수 Y 라 두고 이에 대한 월별, 분기별 평균의 추정식을 살펴보자.

(1) 월별 추정식

$$\begin{aligned} {}_s Y_{qgh} &= \sum_i {}_s Y_{qghi} \\ &= \sum_i ({}_s M_{qghi} \cdot {}_s Y_{qghi}) \end{aligned}$$

여기에서

$$\begin{aligned} {}_s M_{qghi} &= {}_s X_{qghi} / {}_s X_{qgh} , \\ {}_s Y_{qghi} &= \sum_j {}_s Y_{qghij} , \\ {}_s X_{qgh} &= \sum_j {}_s X_{qghij} , \\ Y_{qgh} &= \sum_s {}_s Y_{qgh} \end{aligned}$$

로 주어진다. 단,

Y = 특성 Y 를 갖는 추정 인구,

Y = 표본에서 특성 Y 를 갖는 것으로 조사된 총인구,

M = 승수,

X = 15세 이상 추계인구,

X = 표본에서 조사된 15세 이상 총인구,

s = 성별을 나타내는 첨자($s=1, 2$),

q = 분기를 나타내는 첨자($q=1, 2, 3, 4$),

g = 월을 나타내는 첨자($g=1, 2, \dots, 12$),

h = 시도를 나타내는 첨자($h=1, 2, \dots, 16$),

i = 동부 및 읍면부를 나타내는 첨자($i=1, 2$),

j = 표본 조사구를 나타내는 첨자

(2) 분기별 평균 추정

특성 Y 를 갖는 인구의 성별 및 총계의 분기별 월평균 추정치는 다음

과 같이 산출한다.

$${}_s Y_{qh} = \sum_g {}_s Y_{qgh} / 3 ,$$

$$Y_{qh} = \sum_g Y_{qgh} / 3$$

(3) 추정 오차의 계산

추정치에 대한 추정오차의 산출 공식은 다음과 같다.

① ${}_s Y_{qg}$ 에 대한 표준오차

$$SE({}_s Y_{qg}) = \sqrt{Var({}_s Y_{qg})} ,$$

$$Var({}_s Y_{qg}) = \sum_h Var({}_s Y_{qgh})$$

$$= \sum_h \sum_i Var({}_s Y_{qghi})$$

$$= \sum_h \left(\sum_i ({}_s M_{qghi})^2 \cdot \phi_{hi} \cdot \sum_j ({}_s Z_{qghij})^2 \right)$$

위의 공식에서

$$\phi_{hi} = [(1 - f_{hi}) / n_{hi}] / [2(n_{hi} - 1)] ,$$

$$f_{hi} = n_{hi} / 10N_{hi} ,$$

$${}_s Z_{qghij} = d {}_s Y_{qghij} - {}_s \phi_{qghi} \cdot d {}_s X_{qghij} ,$$

$$d {}_s Y_{qghij} = {}_s Y_{qghij} - {}_s Y_{qghij+1} ,$$

$$d {}_s X_{qghij} = {}_s X_{qghij} - {}_s X_{qghij+1} ,$$

$${}_s \phi_{qghi} = {}_s Y_{qghi} / {}_s X_{qghi} ,$$

n = 표본조사구수,

N = 총추출단위 조사구수

② Y_{qg} 에 대한 표준오차

$$SE(Y_{qg}) = \sqrt{Var(Y_{qg})},$$

$$\begin{aligned} Var(Y_{qg}) &= \sum_h \sum_i Var(Y_{qghi}) \\ &= \sum_h \sum_i \left(\sum_s Var(s Y_{qghi}) + 2Cov({}_1 Y_{qghi}, {}_2 Y_{qghi}) \right) \end{aligned}$$

$$Cov({}_1 Y_{qghi}, {}_2 Y_{qghi}) = {}_1 M_{qghi} \cdot {}_2 M_{qghi} \cdot \phi_{hi} \cdot \sum_j {}_1 Z_{qghij} \cdot {}_2 Z_{qghij}$$

2004년 1월부터 시행되는 연동표본가구 관리체계 하에서 모수 추정은 전월과 금월간의 변동을 고려하여 가중값을 부여하는 다음과 같은 형식의 복합 추정량을 이용해야 할 것이다.

$$Y_t = (1 - k) \hat{Y}_t + k(Y_{t-1} + \Delta_t) + A\hat{\beta}_t$$

여기에서 Y_t 은 해당 월의 추정값이고 \hat{Y}_t 는 해당 월의 직접 추정값이며 Y_{t-1} 는 전 월의 추정값이다. Δ_t 는 두 달간의 변동분이고 $\hat{\beta}_t$ 는 표본교체의 영향을 나타내며 k, A 는 상수이다. 위에 주어진 복합 추정량의 분산 추정값은 잭나이프 추정방법을 적용하면 보다 효율적으로 계산할 수 있다.

2.2 표본관리 및 오차관리

경제활동인구조사에서 표본가구들은 5년 동안 지속적인 조사가 이루어져 표본가구의 부담이 커져서 비표본오차를 유발할 수 있는 원인이 될 수 있기 때문에 연동표본 관리체계를 적용하게 되었지만 표본가구의 교체로 발생하는 시계열의 이탈과 과도한 조사업무를 최소화하는 방안을 마련해야 할 것이다. 새로 교체되는 표본가구와 구 표본가구간의 차이가 크지 않도록 표본관리에 신경을 많이 써야 할 것이다. 교체할 표본가구를 사전

에 선정해야만 시계열 유지에 도움을 줄 수 있을 것이다. 매월 $\frac{1}{36}$ 씩 표본가구를 교체한다면, 전월대비 $\frac{35}{36}$ 의 표본가구는 동일한 가구이고, 전년 동월대비 $\frac{24}{36}$ 의 표본가구가 동일한 것이다. 그러나 표본가구의 이동이 없을 경우에는 위의 언급이 맞지만 이사 등으로 표본가구가 변경된 경우에는 교체된 것으로 보지 않고 계속 조사된 것으로 간주함으로써 비표본오차가 발생할 가능성이 있다.

고정표본가구 관리체계 하에서 표본가구들은 약 5년간 고정 관리되므로 조사 초기에 조사자와 피조사자의 Rapport가 형성되면 5년간 별다른 어려움 없이 조사업무가 수행될 수 있지만 2004년부터 적용되는 연동표본가구 관리체계에서는 $\frac{1}{36}$ 씩의 표본가구가 매월 교체되므로 조사원의 업무가 더 많아지게 될 뿐 아니라 개인의 사생활 보호와 이기주의적인 의식의 팽배로 조사를 거부하는 사례가 많아지게 되므로 정해진 규모의 표본가구 수를 채우지 못하게 됨으로써 경찰특성에 대한 추정값의 정도가 낮아지게 되며 이에 대한 대비 방안도 검토해야 할 것이다.

조사원의 업무부담을 줄일 수 있는 조사방법을 연구하여 조사원이 새로 교체되는 표본가구의 관리에 더 많은 노력을 기울일 수 있도록 하는 것이 바람직하다. 캐나다와 미국의 경우에는 새로 편입되어 오는 가구에 대해서는 방문 면접조사를 실시하고 어느 정도 Rapport가 형성되면 전화조사를 실시함으로써 조사원의 업무를 경감시켜주고 있다. 노트북을 휴대하여 조사한 후에 별도의 시간을 내어 자료를 입력하고 전송하는 노력을 줄일 수 있는 체계적이고 경험적인 연구가 필요하다. 본 연구가 비표본오차 관리에 관한 내용이 아니기 때문에 오차관리에 대해서는 개괄적인 내용만을 언급하였으나 별도의 연구가 필요할 것이다.

2.3 소지역(시군구) 추정 필요성

사회가 글로벌 정보화로 발전하면서 보다 정확하고 시의적절한 통계의 필요성은 더욱 강조되고 있으며 또한 다양한 종류의 통계 생산이 요구되고는 있지만 조사환경은 점점 어려워지고 있는 실정이다. 조사에만 의존하던 통계작성 기법은 개선되어야 하고 국가행정 전산망이 활성화되면서 행정업무나 보고자료를 통해서 얻어진 정보를 통계화하여 조사에 의해서 생산되는 통계와 상호 보완하여 통계의 정확성을 높일 수 있는 방법이 강구되어야 할 것이다. 특히 지방자치행정의 기본 단위인 시군구에 대한 다양한 통계의 수요가 많아지고 있고, 경제활동 관련 통계는 개인 뿐 아니라 사회 또는 정치적으로도 중요한 관심사가 되고 있기 때문에 어느 정도의 신뢰도와 정확성을 갖춘 시군구의 실업통계를 생산할 수 있는 연구는 매우 중요하고 시급한 국가적인 연구과제이다.

지방자치 단체장들이 지역주민들의 직접 선거에 의해 선출되기 때문에 시군구의 단체장들은 시군구의 특성을 정확하게 파악하는 것도 매우 중요하다. 특히 해당 시군구의 고용상태와 경제활동의 변동에 대해서는 더욱 정확하게 파악하고 개선방안을 마련해야 하므로 시군구단위의 실업 관련 통계는 정확하게 시의적절한 시기에 작성되고 공표되어야 할 것이다.

지역의 균형적인 발전을 위해서는 지역의 특성에 알맞고 경제적으로 경쟁력을 가질 수 있는 장기적인 지방행정을 기획하고 추진해야 하는데 여기에 필요한 것이 지역통계이며 이 중에서도 시군구의 고용상태를 파악할 수 있는 경제활동인구 관련 통계는 매월 또는 분기별 혹은 반기별로는 작성 공표하여 투명하고 효율성 있는 지역행정이 이루어져야 할 것이다.

특·광역시와 도 단위에서도 “성별-연령대별” 또는 “성별-교육정도별” 등의 세분화된 통계가 작성되어야 효율적인 고용정책을 수립하여 책임 있

는 행정을 펼 수 있을 것이다. 특히 “성별-연령대별” 또는 “성별-교육정도별”의 경제활동인구 관련 통계는 경제적이고 효율적인 고용증진 정책수립과 행정에도 도움을 줄 수 있고, 지역의 균형발전과 해당 시도에 가장 적합한 발전계획을 수립하여 추진함으로써 지역 재정력을 제고하는데 도움이 될 수 있을 것이다. 좀 더 다양하고 활용성이 큰 통계를 생산할 수 있는 소지역 추정법의 연구는 경제적인 측면에서 적극적으로 추진되어야 할 과제이다.

제3장 소지역 추정법에 의한 노동통계 작성사례

3.1 미 국

3.1.1 서 론

1950년대에 미 노동성 고용훈련 행정국에서는 각 주(state)들에 대한 실업률 추정값들을 비교할 수 있도록 실업관련 통계 추정기법을 개발하여 소책자로 발간하였다. 이 소책자를 근거로 하여 1950년대 후반에는 소규모 표본조사를 통한 연속적인 주 단위 실업통계가 작성되었다. 1950년대의 실업통계는 주로 실업보험(UI:unemployment insurance) 자료를 이용하여 작성되었다.

1972년 미 노동 통계국에서 주 단위에 대해서 이용 가능한 노동력, 취업자 및 실업자 추정에 대한 방법과 개념을 연구하기 시작하였고, 1973년에는 통계국이 주관이 되어, 경상인구조사(CPS:Current Population Survey)의 개념, 정의, 추정과 이전의 소책자방법을 결합하여 주 단위와 주의 세부단위까지 노동력 관련 통계를 추정할 수 있는 기법을 개발하였다.

1976년 이후에는 모든 주 단위 실업 관련 통계의 추정값에 대한 신뢰도를 높이기 위해서 각 주별로 표본 가구 수를 몇 배씩 증가시켰으며, 이후부터 CPS 자료를 이용한 노동력 관련 추정값을 공식적으로 발표하기 위해 실업 관련 추정값에 대한 변동계수의 최대 허용기준을 설정하였다(실업률을 6%라고 가정했을 때, 실업 관련 통계에 대한 변동계수의 최대 허용값을 10%로 함). 1978년부터 규모가 큰 10개주(캘리포니아, 플로리다, 일리노이스, 메사츄셀, 미시간, 뉴저지, 뉴욕, 오하이오, 펜실베이니아, 텍사스)와 2개 지역(로스앤젤레스, 뉴욕시)의 노동력 관련 통계는 CPS자료만으로 추정된 결과를 공식 통계로 사용토록 하였다.

UI 신청자료의 데이터베이스를 지속적으로 유지하고 개선하였으며, 1976년과 1978년 사이에 걸쳐 UI 신청자료를 모든 주에 대해서 표준화하고, CPS조사의 조사 주 간(매월 12일을 포함한 주)에 실업자로 인정받는 UI 신청자는 자동으로 데이터베이스에 등록되도록 하는 데이터베이스 관

리체계를 개발하여 CPS자료와 연계한 추정법을 개발하였다.

1985년에는 1980년 센서스 자료를 근거로 하여 주 단위들에 대한 CPS 표본설계를 완성하였으며, 이 때 추가적으로 노스캐롤라이나 주를 CPS 자료에서 노동력 관련 통계를 직접 추정하여 공표 하는 주로 포함시켰고, 표본크기가 충분한 총 11개 대규모 주의 실업 관련 통계에 대한 목표변동계수를 8%로 낮추었다. 또한 나머지 39개 주의 연평균 실업통계에 대한 목표변동계수를 8%로 정하였으며, 이 때 실업률의 참값은 6%인 것으로 가정하였다.

표본크기가 충분한 11개 주 이외의 39개 주와 주 내의 부차관심영역들에 대한 공식적인 월별 추정값들은 1989년 이전까지는 노동성에서 개발한 소책자 추정방법에 의해 계산하였으나, 1989년부터 시계열 모형을 이용한 추정기법을 표준 추정법으로 채택하였다. 1992년에는 추정값들의 시계열들에 대해 계절요인 보정을 적용하였고 1994년에는 1990년 센서스 자료를 이용하여 CPS를 재 설계하였으며, 또한 모든 조사에서 컴퓨터 보조면접을 실시할 수 있도록 설문지를 재구성하였다. 새로 개편된 CPS의 표본설계는 1995년 중반까지 단계적으로 도입하였다.

CPS는 매월 미국 내의 노동력의 동향과 이와 관련된 세부항목들(연령대별, 성별, 인종별 등)에 대한 특성을 파악하기 위해 실시되는 경상인구 조사으로써 총화 다단확률추출에 의해 추출된 가구단위들에 대해 조사가 이루어진다. CPS의 최종추출단위는 가구조사 단위들로써 약 50,000 조사가 구들에 대해 조사가 실시되며, 조사가구들은 총 8개의 패널로 구성되어 표본으로 관리된다. CPS의 표본교체방식은 4-8-4체계를 따른다. 표본으로 추출된 최종 추출단위(가구단위)들은 4개월 간 연속조사가 진행되며, 이후 8개월 간 조사에서 제외되었다가 다시 4개월 간 연속조사가 진행된 후 표본목록에서 영구히 삭제된다.

매달 8개 패널의 표본가구에 거주하는 만 15세 이상의 사람들을 대상으로 노동력조사가 이루어지며, 노동력 관련 최종 추정값들은 여러 단계의 추정과정을 통해 작성된다. CPS 자료로부터 전국 단위와 주 단위에

대한 노동력 관련 추정값들을 작성하기 위한 개략적인 추정과정은 다음과 같다.

- (a) CPS 가중치를 이용하여 노동력 관련 추정값들의 불편성조정
- (b) 무응답에 대한 보정
- (c) 1차 추출단위들(PSUs)의 분산을 줄이기 위한 일단계 비보정 (First-stage ratio adjustment)
- (d) CPS 추정값들의 분산을 줄이기 위한 이단계 비보정(Second-stage ratio adjustment)
- (e) 추정값들의 분산을 줄이기 위해 전 달의 조사자료를 이용하는 복합 추정
- (f) 주요한 노동력 통계에 대한 계절요인 조정(Seasonal adjustment)

3.1.2 불편성 조정

대부분의 표본조사에서와 마찬가지로 CPS에서도 무응답이 발생한다. 미국 CPS에서의 단위 가구들에 대한 월별 무응답은 평균적으로 6~7%정도이다. 이러한 무응답이 편향의 잠재적인 요인으로 작용하게 된다. 이 밖에 표본선택 과정에서 발생하는 단위 가구들에 대한 누락이라든가 면접자에 의해 가구 단위 또는 가구 내의 개인 조사단위가 누락될 경우 등이 편향의 요인으로 작용될 수 있다. 이러한 편향의 잠재적인 가능성을 최소화하기 위해 추가적으로 무응답에 대한 보정이 추정과정에 포함된다.

표본 내에 있는 모든 단위들이 동일한 추출확률을 갖는다면 이러한 표본은 자체가중치(Self-weighting)를 갖는 표본으로 분류되며(동일한 추출확률을 갖지 않을 경우는 비자가중치(Nonself-weighting)를 가짐), 불편추정량은 표본총계에 추출확률의 역수를 곱하여 계산된다. CPS에서 대부분의 주(state) 지역 표본들은 자체가중치를 갖는 지역들에 해당된다.

(1) 기본 가중치

CPS는 총 792개 PSU(Primary sampling units) 지역(층)들로 구성되며, 이들은 432개의 자체가중치를 갖는 SR(Self-representing) 지역(층)들과 360개의 비자체가중치를 갖는 NSR(Nonself-representing) 지역(층)들로 구분된다. CPS에서 주별 SR 및 NSR 지역(층)들의 수 및 추정인구는 다음 <표3.1>에 주어졌다. 여기에서 추정인구는 1990년 센서스에 근거하여 추정된 결과이다.

<표3.1> 각 주별 층의 수 및 추정인구

State	SR 지역		NSR 지역		전체 추출간격
	층의 수	추정인구	층의 수	추정인구	
Alabama	4	1,441,118	10	1,604,901	2,298
Alaska	5	287,192	5	91,906	336
Arizona	3	2,188,880	4	544,402	2,016
Arkansas	4	612,369	13	1,151,315	1,316
California	21	20,866,697	5	1,405,500	2,700
Los Angeles	1	6,672,035	0	0	1,691
Remainder	20	14,194,662	5	1,405,500	3,132
Colorado	5	1,839,521	5	629,640	1,992
Connecticut	20	2,561,010	0	0	2,307
Delaware	3	509,330	0	0	505
District of Columbia	1	486,083	0	0	356
Florida	20	9,096,732	8	1,073,818	2,176
Georgia	9	2,783,587	9	2,038,631	3,077
Hawaii	3	778,364	1	49,720	769
Idaho	8	416,284	8	302,256	590
Illinois	11	6,806,874	12	1,821,548	1,810
Indiana	9	2,215,745	8	1,949,996	3,132
Iowa	5	710,910	11	1,372,871	1,582
Kansas	3	920,819	9	906,626	1,423
Kentucky	9	1,230,062	9	1,546,770	2,089
Louisiana	7	1,638,220	9	1,403,969	2,143
Maine	8	685,937	4	247,682	838
Maryland	4	3,310,546	2	352,766	3,061
Massachusetts	31	4,719,188	0	0	954
Michigan	12	5,643,817	11	1,345,074	1,396
Minnesota	4	2,137,810	7	1,119,797	2,437
Mississippi	6	540,351	14	1,334,898	1,433

Missouri	7	2,394,137	6	1,457,616	3,132
Montana	8	366,320	8	221,287	431
Nebraska	2	557,203	9	608,513	910
Nevada	4	819,424	2	98,933	961
New Hampshire	13	846,029	0	0	857
New Jersey	11	6,023,359	0	0	1,221
New Mexico	6	685,543	8	410,929	867
New York	13	12,485,029	7	1,414,548	1,709
New York City	1	5,721,495	0	0	1,159
Remainder	12	6,763,534	7	1,414,548	2,093
North Carolina	14	3,089,176	23	1,973,074	1,095
North Dakota	4	232,865	9	235,364	363
Ohio	13	6,238,600	13	1,958,138	1,653
Oklahoma	2	1,244,920	11	1,092,513	1,548
Oregon	4	1,396,261	6	761,794	1,904
Pennsylvania	14	7,704,963	11	1,507,946	1,757
Rhode Island	5	784,090	0	0	687
South Carolina	7	1,434,621	7	1,161,298	2,291
South Dakota	5	211,146	11	291,546	376
Tennessee	8	2,502,671	6	1,219,915	3,016
Texas	22	8,779,997	20	3,613,170	2,658
Utah	2	888,524	3	251,746	958
Vermont	11	428,263	0	0	410
Virginia	12	3,093,947	6	1,612,667	3,084
Washington	5	2,437,454	6	1,216,557	2,999
West Virginia	10	803,797	9	581,544	896
Wisconsin	7	1,773,109	9	1,889,141	2,638
Wyoming	8	227,401	6	98,321	253
Total	432	141,876,295	360	45,970,646	2,060

<표3.1>에 주어진 해당 주 별 표본 추출간격을 CPS 기본가중치로 이용한다. 같은 주 내에 있는 대부분의 표본조사 단위들은 동일한 추출확률을 가지나 New York과 California 내에서는 예외적으로 서로 다른 추출확률이 적용된다. 추정과정의 첫 번째 단계에서는 단순 집계된 가구 조사 단위들에 대해서 위의 기본가중치가 곱해진다. 인구의 증가에 따라 가구 단위들이 증가할 경우 원래의 표본추출간격은 조사 비용과 정확도의 측면을 고려하여 정기적으로 보정된다.

(2) 특별 가중치

CPS에서 최종 추출단위들은 보통 4가구로 구성된다. 경우에 따라서는 최종 추출단위 내에 4가구 이상의 조사가구들이 포함되어 있는 경우가 있을 수 있다. 예를 들면 다음과 같은 예이다. CPS에서 하나의 최종 추출단위 지역에는 4개의 조사가구가 배정되는 것이 일반적이지만 조사자가 면접 당시 하나의 최종 추출단위 지역에 36개 가구단위가 존재하였다고 가정해 보자. 조사자의 업무 부담을 줄이고자 3가구 당 하나의 가구를 부차적으로 추출하여 총 12개 가구를 조사했을 경우 CPS에서는 특별가중치를 3으로 산정하여 해당 지역의 기본가중치에 곱하여 추정값들이 불편성을 만족하도록 보정해 준다. CPS에서는 추정값의 편향과 분산을 고려하여 특별가중치가 4를 넘지 않도록 명시하고 있다.

3.1.3 무응답에 대한 보정

CPS에서는 무응답을 크게 항목무응답(Item nonresponse)과 단위무응답(Unit nonresponse)의 두 가지로 구분한다. 여기에서 단위무응답은 해당 표본단위로부터 조사자료를 수집하지 못한 경우를 의미한다. 예를 들면, 피면접자의 응답 거부 또는 피면접자의 부재 등의 사유로 조사자료를 수집하지 못한 경우 등이 여기에 해당된다. CPS에서는 이러한 단위무응답을 Type A 무응답이라 부르며 매월 약 4~5%의 Type A 무응답이 발생한다. 최근에 들어서는 Type A 무응답의 비율이 6~7%정도로 다소 증가하는 경향을 보인다. CPS에서는 단위무응답 요인들을 고려하여 가중치에 대한 보정을 실시한다.

단위무응답에 대한 보정은 다음과 같이 이루어진다. 우선 해당 주에서 광역통계구역(metropolitan statistical area: MSA)과 크기가 비슷한 1차 추출단위(PSU)들을 한데 묶어 MSA 집락으로 구분하고, 그렇지 못한 PSU들은 non-MSA 집락으로 구분한 후, 각 집락 내에서는 두 개의 무응답보정 셀을 형성한다. MSA 집락은 “도시 중심지역”과 “기타 지역”으로, non-MSA 집락은 “도시 지역”과 “농촌 지역”으로 무응답보정 셀이 형성

된다. CPS에서는 총 254개의 무응답보정 셀에 대해서 단위무응답에 대한 보정이 실시되며, 각 무응답보정 셀에 대한 무응답보정값 F_{ij} 는 다음과 같이 산정된다.

$$F_{ij} = \frac{Z_{ij} + N_{ij}}{Z_{ij}},$$

여기에서 “ $Z_{ij}=i$ 집락의 j 셀에서 조사된 가구단위들의 가중총계”, “ $N_{ij}=i$ 집락의 j 셀에서 조사된 가구단위들의 Type A 무응답 가중총계”를 나타낸다.

조사된 개인단위들은 위에서 산정된 무응답보정값 F_{ij} 가 곱해져서 단위무응답에 대한 보정이 이루어진다. 참고적으로 CPS에서는 F_{ij} 가 2이상인 경우 또는 해당 셀에서 조사된 무응답 조사가구가 50보다 작을 경우에는 예외적인 보정절차가 시행된다. 무응답보정이 이루어진 후 각 개인 조사단위들에 대한 가중치는 “기본가중치×특별가중치×무응답보정값”이 적용된다.

3.1.4 비 보정(Ratio Adjustment)

(1) 일단계 비보정

일차추출단위(PSU)에서 발생하는 주 수준(state level) 추정값들의 분산을 줄이기 위한 목적으로 일단계 비보정이 실시된다. 일단계 비보정값들은 1990년 센서스 자료에 근거하여 산정되며, 비자체가중치를 갖는 NSR 지역(층)의 PSU들에 대해서 적용된다. NSR 지역(층)의 PSU들을 포함하고 있는 각 주별로 두 가지의 인종별 범주(Black, non-Black)에 대해서 일단계 비보정값들이 계산되며, 산정공식은 다음과 같이 주어진다.

$$FS_{sj} = \frac{\sum_{i=1}^n C_{sij}}{\sum_{k=1}^m \left(\frac{1}{\pi_{sk}} \right) C_{skj}},$$

여기에서

$FS_{sj} = s$ 주의 j 인종 셀에 대한 일단계 비보정값($j = \text{Black, non-Black}$),
 $C_{stj} = s$ 주의 j 인종 셀에서 NSR 지역(층)의 전체 PSU 중 i PSU에 대
 한 만 15세 이상의 1990년 센서스 인구,
 $C_{stk} = s$ 주의 j 인종 셀에서 NSR 지역(층)의 표본 PSU 중 k PSU에
 대한 만 15세 이상의 1990년 센서스 인구,
 $\pi_{sk} = s$ 주에서 k PSU에 대한 1990년 추출확률,
 $n = s$ 주에서 NSR 지역(층)에 있는 PSU들의 전체 개수,
 $m = s$ 주에서 NSR 지역(층)의 표본 PSU들의 개수.

만약, s 주에서 일단계 비보정값이 1.3보다 크거나 $1/1.3$ 보다 작은 경
 우, NSR지역(층)의 표본 PSU들이 4개보다 작은 경우에 대해서는 인종
 셀의 범주를 Black 또는 non-Black으로 구분하지 않고 일단계 비보정값
 을 공통적으로 1을 부여한다. 자체가중치를 갖는 SR 지역(층)에 대해서도
 마찬가지로 일단계 비보정값은 1이 부여된다. CPS의 일단계 비보정값들
 은 다음 <표3.2>에 주어졌다. 일단계 비보정이 이루어진 후 각 개인 조사
 단위들에 대한 가중치는 “기본가중치×특별가중치×무응답보정값×일단계
 비보정값”이 적용된다.

<표3.2> 각 주별 일단계 비보정값(1999년 CPS 기준)

State	Black	non-Black
Alabama	0.92986976	1.02360321
Alaska	**1.00000000	**1.00000000
Arizona	**1.00000000	**1.00000000
Arkansas	1.03854268	0.99625041
California	0.92824011	1.01550280
Colorado	**1.00000000	**1.00000000
Connecticut	*1.00000000	*1.00000000
Delaware	*1.00000000	*1.00000000
District of Columbia	*1.00000000	*1.00000000
Florida	1.07779736	1.00025192
Georgia	1.06981965	0.98237807

Hawaii	**1.00000000	**1.00000000
Idaho	**1.00000000	**1.00000000
Illinois	1.01947743	1.00254363
Indiana	1.16715920	0.99747055
Iowa	**1.00000000	**1.00000000
Kansas	**1.00000000	**1.00000000
Kentucky	1.09352656	0.99897341
Louisiana	1.04956759	0.98344772
Maine	**1.00000000	**1.00000000
Maryland	**1.00000000	**1.00000000
Massachusetts	*1.00000000	*1.00000000
Michigan	0.92441097	0.99798724
Minnesota	**1.00000000	**1.00000000
Mississippi	0.98243024	1.00997154
Missouri	**1.00000000	**1.00000000
Montana	**1.00000000	**1.00000000
Nebraska	**1.00000000	**1.00000000
Nevada	**1.00000000	**1.00000000
New Hampshire	*1.00000000	*1.00000000
New Jersey	*1.00000000	*1.00000000
New Mexico	**1.00000000	**1.00000000
New York	0.83647167	1.00163779
North Carolina	1.07378643	0.98057928
North Dakota	**1.00000000	**1.00000000
Ohio	0.97362223	1.00000085
Oklahoma	1.01775196	1.00400963
Oregon	**1.00000000	**1.00000000
Pennsylvania	1.17560284	0.99367587
Rhode Island	*1.00000000	*1.00000000
South Carolina	0.93971915	1.05454366
South Dakota	**1.00000000	**1.00000000
Tennessee	1.08638935	0.98680199
Texas	1.23277658	0.97475648
Utah	**1.00000000	**1.00000000
Vermont	*1.00000000	*1.00000000
Virginia	**1.00000000	**1.00000000
Washington	**1.00000000	**1.00000000
West Virginia	1.22587622	0.99959222
Wisconsin	**1.00000000	**1.00000000
Wyoming	**1.00000000	**1.00000000

* 자체가중치를 갖는 SR 지역(총) (총 8개 주)

** 인종 셀의 범주(Black, non-Black)가 합쳐진 지역 (총 23개 주)

(2) 이단계 비보정

이단계 비보정은 노동력과 관련된 세부항목과 관심영역 및 부차관심 영역들에 대한 표본 추정값들의 추정오차를 줄이기 위한 목적으로 실시되며, 각 패널 내의 표본 가중치들에 대해서 보정이 이루어진다. CPS에서는 이단계 비보정 시 매월 서로 독립적으로 기록 유지되는 다음과 같은 세 가지 대조그룹들에 대한 정보들을 이용한다.

- 50개 주와 콜롬비아 구에 대한 만 15세 이상의 추계인구
- 라틴 아메리카 계의 14개 성별-연령대별 추계인구, 비 라틴 아메리카 계의 5개 성별-연령대별 추계인구(세부범주는 <표3.3>을참조)
- 백인에 대한 66개 성별-연령대별 추계인구, 흑인에 대한 42개 성별-연령대별 추계인구, 기타 10개 성별-연령대별 추계인구(세부범주는 <표3.4>를 참조)

<표3.3> 라틴아메리카계(Hispanic)/비 라틴아메리카계 (non-Hispanic)에 대한 성별-연령대별 범주

Ages	Hispanic	
	Male	Female
0-5		
6-13		
*14		
*15		
16-19		
20-29		
30-49		
50+		

Ages	non-Hispanic
*0-5	
*6-13	
*14	
*15	
*16+	

* 표시는 성별 구분이 없는 범주를 나타냄

<표3.4> 백인(White)/흑인(Black)/기타에 대한 성별-연령대별 범주

Ages	White male	White female
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10-11		
12-13		
14		
15		
16		
17		
18		
19		
20-24		
25-26		
27-29		
30-34		
35-39		
40-44		
45-49		
50-54		
55-59		
60-62		
63-64		
65-67		
68-69		
70-74		
75+		

Ages	Black male	Black female
0-1		
2-3		
4-5		
6-7		
8-9		
10-11		
12-13		
14		
15		
16-17		
18-19		
20-24		
25-29		
30-34		
35-39		
40-44		
45-49		
50-54		
60-64		
65+		

Ages	Other male	Other female
0-5		
6-13		
*14		
*15		
16-44		
45+		

*는 성별 구분이 없는 범주를 나타내며, 기타 인종은 아메리카 인디언 계, 에스키모 계, 알류트 계, 아시아 계, 태평양 인근 섬 주민들이 포함됨.

이단계 비보정은 월 단위 표본을 구성하는 8개의 패널들 각각에 대해서 개별적으로 이루어진다. 한편, 각 대조그룹에서 패널별 추계인구와 표본 추정값 간의 비보정(ratio adjustment)값들은 대조그룹들에 따라서 약간의 차이가 발생한다. CPS에서는 세 가지 대조그룹들에 대한 정보를 동

시에 비보정값 산정에 포함시킬 수 있는 레이킹 비 추정량(raking ratio estimator)을 개발하여 이단계 비보정값들을 산정하고 있다. 레이킹 비추정량에 대해서는 Ireland and Kullback(1968)에 보다 상세히 소개되어 있다.

이단계 비보정이 이루어진 후 각 개인 조사단위들에 대한 가중치는 “기본가중치×특별가중치×무응답보정값×일단계 비보정값×이단계 비보정값”이 적용된다. CPS에서는 위의 가중치가 적용된 추정값을 FSC(first- and second-stage combined)추정값이라 부른다.

3.1.5 복합 가중치(Composite Weights)

CPS에서 복합가중치 산정은 두 단계에 걸쳐 이루어진다. 우선 인구통계적 특성에 의해 분류된 주요 노동력 범주들에 대한 복합추정값(composite estimates)들이 첫 번째 단계에서 추정된다. 다음으로 각 범주별 개인조사단위에 대한 가중합이 추정된 복합추정값과 일치하도록 일련의 비보정 과정을 통하여 보정된다.

먼저 복합추정값 산정 단계에 대해서 살펴보도록 하자. 일반적으로 복합추정값들은 여러 추정값들의 가중평균의 형식을 취한다. CPS에서의 복합추정값은 다음 두 개의 추정값들을 결합하여 작성된다. 두 추정값들 중 하나는 앞 절에서 언급한 FSC추정값이다. 나머지 하나의 추정값은 바로 전 달의 복합추정값과 전 달과 금번 달의 변화량에 대한 추정값의 항을 결합하여 표현한다. 여기에서 월 변화량에 대한 추정값은 금번 달과 바로 전 달의 CPS 조사표본 중 동일한 표본가구들에 대해 연속적으로 조사되는 약 75%의 조사표본 자료에 근거하여 작성된다(CPS 조사표본의 1개월 간격 중복율은 약 75%임).

t번째 달의 실업자 총계에 대한 복합추정값 Y'_t 의 추정공식은 다음과 같이 주어진다.

$$Y'_t = (1 - K) \hat{Y}_t + K(Y'_{t-1} + \Delta_t) + A \hat{\beta}_t .$$

여기에서

$$\hat{Y}_t = \sum_{i=1}^8 x_{t,i},$$

$$\Delta_t = \frac{4}{3} \sum_{i \in S} (x_{t,i} - x_{t-1,i-1}),$$

$$\hat{\beta}_t = \sum_{i \notin S} x_{t,i} - \frac{1}{3} \sum_{i \in S} x_{t,i},$$

$$i = 1, 2, \dots, 8,$$

$x_{t,i}$ = t 번째 달의 i 표본에서 응답자들에 대한 이단계
비보정 후의 가중합,

$S = \{2, 3, 4, 6, 7, 8\}$: 전 달과 공통으로 조사되는
6개의 패널들의 집합,

$$K = \begin{cases} 0.4, & \text{실업자총계 추정 시} \\ 0.7, & \text{취업자총계 추정 시} \end{cases}$$

$$A = \begin{cases} 0.3, & \text{실업자총계 추정 시} \\ 0.4, & \text{취업자총계 추정 시} \end{cases}$$

위에서 명시된 상수 A 와 K 는 실업자 및 취업자 총계에 대한 월간 변화량의 추정치의 분산을 최소화하는 최적가중치들이다. 상수 K 는 금번 달의 FSC추정치 및 전 달의 복합추정치 Y'_{t-1} 와 금번 달과 전 달의 변화량에 대한 추정치 Δ_t 의 합에 대한 가중치를 결정한다. 여기에서 변화량에 대한 추정치 Δ_t 는 금번 달과 전 달의 공통 조사표본인 6개의 패널의 조사자료에 근거하여 추정된다. 상수 A 는 $\hat{\beta}_t$ 의 가중치를 결정하며, 여기에서 $\hat{\beta}_t$ 는 복합추정량의 분산과 표본 편향을 줄이기 위하여 산정되는 보정항이다(자세한 내용은 Breaux and Ernst(1983)과 Bailar(1975)를 참조).

복합가중치는 주요 노동력 범주인 실업자(UE:unemployed), 취업자(E:employed), 비노동인구(NILF:not in labour force)의 세가지 범주에 대해서 각각 작성되며 실업자에 대한 복합가중치 산정절차를 설명하면 다음과 같다.

(i) j 번째 주(50개 주와 1개의 콜롬비아 구로 구성된 총 51개의 셀)에 대해서 실업자 총계에 대한 복합추정값 $comp(UE_j)$ 를 계산한다. 같은 방법으로 전국 9개의 연령대별-성별-인종별 셀(라틴 아메리카 계에 대한 8개의 연령대별-성별 셀, 비 라틴 아메리카 계에 대한 1개의 연령대별-성별 셀들로 구성)들과 66개 연령대별-성별-인종별 셀(백인에 대한 38개의 연령대별-성별 셀, 흑인에 대한 24개의 연령대별-성별 셀, 기타 4개의 연령대별-성별 셀들로 구성)들에 대한 복합추정값들을 계산한다(연령대별-성별-인종별 셀에 대해서는 <표3.3>과 <표3.4>를 참조). 상수 K 와 A 는 실업자 총계 추정에서는 모든 범주들에 대해서 $K=0.4$, $A=0.3$ 이 적용된다.

(ii) 조사된 표본자료를 각 주별로 분류하고, j 번째 주 내에서 실업자 총계에 대한 단순추정값 $simp(UE_j)$ 을 계산한다. 여기에서 $simp(UE_j)$ 는 이단계 비보정 후의 가중치가 적용된 추정값이다.

(iii) j 번째 주 내에서 실업자에 해당하는 각 개인 조사단위가 갖는 가중치에 다음과 같은 비보정 가중치를 곱한다.

$$\frac{comp(UE_j)}{simp(UE_j)}$$

(iv) 표본조사자료를 연령대별-성별-인종별로 교차분류한 후, 각 셀 내에서 실업자 총계에 대한 단순추정값을 계산한다.

(v) 각각의 연령대별-성별-인종별 셀 내에서 실업자에 해당하는 개인 조사단위들은 절차(iii)과 같은 방법으로 가중치 보정이 이루어진다.

(vi) 절차(iv)와 절차(v)의 과정이 연령대별-성별-인종별 셀들에 대해서 반복된다.

(vii) 절차(ii)에서 절차(vi)까지의 과정이 5회 더 반복된다. 따라서의 총 6회의 반복이 이루어진다.

실업자 및 비노동력 인구에 대해서도 위와 같은 방법이 동일하게 적용

되며, 각 개인조사단위들은 위에서 언급된 복합가중치 적용절차가 끝난 후의 가중치를 최종가중치로 적용받게 된다.

3.1.6 계절요인 보정(seasonal adjustment)

시계열은 시간에 따라 측정 가능한 관측치 또는 추정값들의 계열을 말한다. CPS 자료에 기반을 두고 있는 많은 월별 노동력 통계 자료는 시계열 자료 분석방법에 의해 계절요인이 보정되어 노동력 통계의 시의성 및 정확도를 배가시키고 있다.

계절요인의 보정은 날씨, 휴일, 학기 일정 등과 같은 일상적인 계절 요인 효과를 시계열로부터 측정하여 노동력 지표들을 조정하기 위한 목적으로 실시한다. 실업 및 취업 등과 같은 다양한 추정치들의 월별 변화 양상도 계절요인으로 보다 쉽게 설명될 수 있으며, 주요 경제지표들에 대한 장기순환 및 추이성에 대한 조정도 계절요인의 분석에 의해 이루어질 수 있다. 예를 들어 5월과 6월의 실업 및 취업자 총계에 대한 추정값들을 고려해 보자. 이 경우 6월의 실업 및 취업자 총계는 5월보다는 현격한 증가가 발생할 수 있으며, 학기가 끝나는 5월말을 기준으로 6월에는 노동력인구에 학생들이 유입되기 때문에 경제지표에 대한 장기적인 순환과 추이성에는 영향을 미쳤다고 보기는 어렵다. 따라서 이러한 시의성이 반영되어 6월의 시계열은 5월의 자료에 근거하여 유사하게 보정된다.

계절요인 보정에 대한 연구는 오랜 기간 동안 꾸준히 진행되어오고 있다. 노동력 관련 시계열들의 계절요인 보정을 위해 1980년 이후 지금까지 X-11 ARIMA 프로그램이 이용되고 있다. 이 프로그램은 1960년대에 미 센서스 국에 의해 개발된 X-11 프로그램을 기반으로 하여 1970년대 말 캐나다 통계청에 의해 보완된 시계열 보정 프로그램이다. 공식적으로 일년에 두 번(6월과 12월) 실업자, 취업자 및 실업률 등과 같은 노동력 관련 시계열들의 계절요인 보정 추정값들이 X-11 ARIMA를 통해 얻어진다.

3.1.7 주(state) 및 주 내의 부차관심영역에 대한 추정

대영역인 51개 주와 주 내의 부차관심영역인 소지역들에 대한 실업관련 추정값들은 각 지역의 경제상황을 나타내는 중요한 지표로 활용된다. 미 노동 통계국에서는 주별 협력 프로그램 하에서 매월 약 6,950개 지역들에 대해서 노동력 및 실업관련 추정값들을 작성하고 있다. 미국 내의 모든 주, 광역시, 구, 인구 25,000이상의 시지역, New England 내의 모든 시지역들과 지방 중심지들이 이러한 추정대상 지역들에 해당된다. 주 및 주 내의 소지역들에 대한 표본의 할당 및 추정결과의 공표뿐만 아니라 이들 지역들의 예산 보조에 대한 적격성 여부는 연방 프로그램에 근거하여 결정된다.

주 및 지방정부는 고용정책 입안과 예산 책정 시 주 및 부차관심영역들에 대한 노동력 관련 추정값들을 활용하며, 이러한 추정값들은 대부분 CPS자료에 근거하여 작성되나 CPS 표본이 충분하지 않은 관심영역들에 대해서는 경상고용통계(CES:current employment statistics) 자료와 실업보험(UI:unemployment insurance) 자료가 추가적으로 추정과정에 활용된다.

주 지역들에 대한 노동력관련 연 평균 추정값들은 CPS자료로부터 직접적으로 산출된다. 또한 충분한 표본을 갖고있는 캘리포니아, 플로리다, 일리노이스, 메사츄셀, 미시간, 뉴저지, 뉴욕, 노스캐롤라이나, 오하이오, 펜실베니아, 텍사스의 11개 주 지역들과 뉴욕시와 로스앤젤레스-롱비치의 2개 지역들에 대해서는 CPS 월 추정값들이 활용된다.

나머지 39개 주들과 콜롬비아 구에 대해서는 CPS 표본이 충분하지 않기 때문에 CES 자료와 UI 자료를 결합시킨 시계열모형으로부터 월 추정값들이 작성된다. 이러한 지역들에 대한 시계열 추정방법을 소개하면 다음과 같이 요약된다.

시점 t 에서 CPS 노동력관련 추정값 $y(t)$ 는 다음과 같은 두개의 독립적인 확률과정들의 합으로 표현된다.

$$y(t) = \theta(t) + e(t). \quad (3.1)$$

여기에서 $\theta(t)$ 는 노동력 참값이고 $e(t)$ 는 표본 오차를 나타낸다. 이 모형을 통해 CPS 복합추정량보다 추정오차가 작은 $\theta(t)$ 를 추정하는 것이 기본적인 목적이다.

(3.1)식에서 시그널 항인 $\theta(t)$ 는 다음과 같은 모형을 통해 추정된다.

$$\theta(t) = \theta'(t) + \eta(t)$$

여기에서

$$\theta'(t) = X(t)\beta(t) + T(t) + S(t) ,$$

$X(t)$ = 기지인 설명변수들의 벡터,

$\beta(t)$ = 랜덤 계수 벡터,

$T(t)$ = 추이 항(trend component),

$S(t)$ = 계절요인 항(seasonal component),

$\eta(t)$ = 노이즈 항(noise componet).

이러한 항들의 각각에 대한 확률과정적인 특성들은 서로 독립적인 정규분포를 따르는 하나이상의 백색 잡음 항(white noise component)들에 의해 결정된다. 각 항들을 j 로 표현할 때 백색 잡음 항들의 분포는 다음과 같이 가정된다.

$$v_j(t) \sim NID(0, \sigma_v^2)$$

여러 개의 항들 중 우선 회귀 항을 살펴보자. 회귀 항을 $M(t)$ 라 놓을 때 이 항은 시간에 따라 변화하는 계수 $\beta(t)$ 와 설명변수들 $X(t)$ 로부터

$$M(t) = X(t)\beta(t)$$

와 같이 주어진다. 위의 회귀 항에 포함된 독립변수들은 관측가능한 변수들이며 CPS 표본 오차와 서로 독립인 UI 자료, CES 자료, 표본 센서스 자료들로부터 선택된다. 회귀계수 $\beta(t)$ 는 다음과 같은 랜덤워크과정으로 모형화된다.

$$\beta(t) = \beta(t-1) + v_\beta(t) ,$$

여기에서

$$v_{\beta}(t) \sim NID(0, \sigma_{\beta}^2)$$

로 주어진다. 한편, 선택된 설명변수들은 CPS 추정값들의 중요한 변동을 모형에 반영하고 있지만 비표본 오차에 대한 변화량은 포함하고 있지 않기 때문에 추가적으로 추이 항과 계절요인 항들이 모형에 포함된다.

추이 항 $T(t)$ 는 다음과 같이 랜덤인 $T(t)$ 와 기울기 $R(t)$ 의 일차근사로 주어진다.

$$T(t) = T(t-1) + R(t-1) + v_T^*(t),$$

$$R(t) = R(t-1) + v_R(t).$$

여기에서

$$v_T^*(t) = \sum_{k=1}^m \delta_k \xi_k(t) + v_T(t),$$

$$\xi_k(t) = \begin{cases} 1 & \text{if } t = t_k \\ 0 & \text{if } t \neq t_k \end{cases}$$

로 주어지며, 계수 δ 는 시점 t 에서의 외부충격을 나타낸다. 교란 항 (disturbance term) $v_T(t)$ 와 $v_R(t)$ 는

$$v_T(t) \sim NID(0, \sigma_T^2),$$

$$v_R(t) \sim NID(0, \sigma_R^2),$$

$$E[v_T^*(t)v_T^*(t)] = 0$$

로 가정된다. 분산이 클수록 추이성은 크게 나타나며, 특히 $v_T(t)$ 의 효과는 상하의 폭을, $v_R(t)$ 는 기울기의 변화를 나타낸다.

계절요인 항 $S(t)$ 는 12달의 도수와 관계가 있는 6개의 삼각함수들의 합으로 주어진다.

$$S(t) = \sum_{j=1}^6 S_j(t).$$

각 항들은

$$S_j(t) = \cos(w_j)S_j(t-1) + \sin(w_j)S_j^*(t-1) + v_s(t) ,$$

$$S_j^*(t) = -\sin(w_j)S_j(t-1) + \cos(w_j)S_j^*(t-1) + v_s^*(t)$$

와 같은 확률변수들의 쌍으로 표현되며, w_j 는 $w_j = 2\pi j/12$ 로 주어진다. 여기에서 v_s 와 v_s^* 는 서로 상관관계가 없는 백색 잡음 확률과정이며 평균 0, 공통분산 σ_s^2 을 갖는 확률과정으로 가정된다.

노이즈 항 $\eta(t)$ 는 위에서 언급된 항들로는 설명될 수 없는 불규칙적인 변동과 정상적인 과정에서 벗어난 일시적인 대규모 변동(또는 이상치)의 합으로 주어진다.

$$\eta(t) = I(t) + O(t)$$

여기에서 $I(t)$ 는 보통 불규칙변동 항(irregular component)으로 불리며, 보통 시그널과 표본오차 항들을 추정한 후에도 여전히 시그널 항에서 불규칙적인 변동이 탐지될 때 고려되는 항이다. 불규칙변동 항 $I(t)$ 는

$$I(t) = v_I(t), \quad v_I(t) \sim N(0, \sigma_I^2)$$

로 주어진다. 만약, $I(t)$ 항의 분산이 0일 경우에는 $I(t)$ 항은 0으로 처리되어 모형에서 제외될 수 있다. 이상치의 항 $O(t)$ 는 관측된 계열에서 탐지되는 일시적인 대규모 변동을 나타내며

$$O(t) = \sum_j \lambda_j \zeta_j(t)$$

로 주어진다. 여기에서

$$\zeta_j(t) = \begin{cases} 1 & \text{if } t=j \\ 0 & \text{otherwise} \end{cases}$$

계수 λ_j 는 시점 j 에서 계열들의 변화량을 나타낸다. $O(t)$ 는 주로 CPS 관측값에 영향을 주는 외부적인 요인에 의해 발생한다. 예를 들면 잘못된 표본가구들의 선정에 의해 관측된 비정상적인 CPS 추정값들은 외부적인 요인에 의해 발생된 이상치로 볼 수 있으며, 이로 인해 발생하는 일시적인 변동은 시계열에 그대로 반영되어 이상치에 대한 인식을 용이하게 해

주며 시그널 항 추정에서 이상치의 양을 줄여주는 보정과정이 이루어진다.

표본오차 $e(t)$ 는 모집단의 값 $\theta(t)$ 와 추정값 $y(t)$ 간의 차

$$e(t) = y(t) - \theta(t)$$

로 정의된다. 여기에서 표본오차 $e(t)$ 는 다음과 같은 특성을 갖는다.

$$E[e(t)] = 0 ,$$

$$\text{Var}[e(t)] = \sigma_{e(t)}^2 ,$$

$$\rho_e(l) = \frac{E\{e(t)e(t-l)\}}{\sigma_{e(t)}^2} .$$

CPS 표본오차의 자기상관적 구조는 CPS 패널 설계와 모집단의 특성에 의해 결정되며, 표본오차의 절대 크기는 표본 재설계, 표본크기의 변화, 노동력 수준들의 변동에 의해 고정되어 있는 값이 아니라 항상 변화하는 값이다. CPS 분산은 시간에 따른 변화가 심하고, 자기상관구조는 이에 비해 매우 안정적이라는 데에 착안하여

$$e(t) = \gamma(t)e^*(t)$$

와 같은 표본오차의 의 형태를 고려할 수 있다. 위 표현식은 $e(t)$ 의 자기상관적인 구조와 이분산적인 구조(heteroscedastic structure)를 파악할 수 있게 해 준다. 분산 팽창계수 $\gamma(t)$ 는 CPS에서 이분산성을 설명하며

$$\gamma(t) = \frac{\sigma_{e(t)}}{\sigma_e}$$

로 정의된다. 여기에서 $\sigma_{e(t)}$ 는 CPS 추정값에 대한 일반화분산함수에 의한 추정값이며, σ_e 는 자기회귀이동평균(ARMA:autoregressive moving average) 확률과정 $e^*(t)$ 의 추정오차이다.

CPS 자기상관 구조는 ARMA 확률과정 모형

$$e^*(t) = \phi^{-1}(L)\theta(L)v_e(t)$$

를 통해 얻는다. 여기에서 L 은 시차 연산자로서 $L^k(X_t) = X_{t-k}$ 로 정의되

며, $\phi(L)$ 과 $\theta(L)$ 은 각각

$$\begin{aligned}\phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p, \\ \theta(L) &= 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q\end{aligned}$$

와 같이 주어진다. 모수 $\phi_1, \phi_2, \dots, \phi_p$ 와 $\theta_1, \theta_2, \dots, \theta_q$ 는 자기상관함수

$$\rho_e(l) = \frac{\theta(L)\theta(L^{-1})}{\phi(L)\phi(L^{-1})}$$

를 통해 표본오차 시차 상관계수(lag correlation)들로부터 추정된다 (Dempster and Hwang, 1990). 여기에서 $\rho_e(-l) = \rho_e(l)$ 가 가정된다.

계수 θ 와 ϕ 는 충격응답 가중치(impulse response weights) $\{g_k\}$ 를 계산 하는데 이용되며, 가중치 $\{g_k\}$ 는 생성함수

$$g(L) = \phi^{-1}(L)\theta(L)$$

를 통해 계산된다. ARMA 확률과정 $e^*(t)$ 의 분산은

$$\sigma_e^2 = \sum_{k=0}^{\infty} g_k^2$$

로 계산된다.

한편, 주 내의 모든 구 지역들과 인구 25,000이상의 시지역들 및 주에 부속되어 있으면서 노동시장에 크게 영향을 미치는 부차관심영역들에 대한 추정값들도 CPS 자료에 CES 자료, UI 자료와 센서스 자료가 추가되어 추정된다.

주 지역들에 대한 추정값들은 X-11 ARIMA 계절요인 보정절차에 따라 일년에 두 번 계절요인에 대한 보정이 이루어진다.

미 센서스 국은 일년에 한번 추계인구에 대한 전체적인 조정을 실시한다. 따라서 모든 주지역들과 콜롬비아 구에 대해서 CPS 노동력 관련 추정값들은 새로운 추계인구를 기준으로 벤치마킹이 이루어진다. Denton 방법(Denton, 1971)으로 불리우는 벤치마킹 절차를 적용하여 모형기반 노동력 관련 추정값들의 연간 평균을 해당되는 CPS 추정값들의 연간 평균과 일치시키는 작업이 이루어진다.

3.1.8 적용 사례

(1) 회귀 모형 추정법

회귀모형을 이용한 추정법은 인근 유사지역의 정보를 이용하여 추정된 합성 추정값이 지역 변동을 충분히 설명하지 못한 점을 보완하고자 실직보험 자료와 구직등록 자료를 이용하여 1970년대 후반부터 시군구의 실업통계 작성에 이용되었던 기법이다. CPS의 1차 추출단위(PSU:primary sampling unit)에 대한 추정치를 종속변수로 사용하고, 다음과 같은 몇 개의 적절한 독립변수를 선택한 회귀모형을 고려한다.

① 다음의 범주를 고려한 합성추정값

- 직업별-성별-인종별 범주
- 결혼여부-연령대별-성별-인종별 범주

② 센서스에서 추정한 실업자 총계 대비 3-4월의 실직보험 가입자에 대한 가입비율(%)

③ 연말 자료에서 공표되는 “70-단계”의 실업률 추정치

5개월 간의 월별 CPS 추정값들의 평균을 종속변수로 사용하기 위해 CPS-PSU와 SMSA를 대응시켰을 때 150개의 SMSA 중 122개의 SMSA가 완전 대응되며 이러한 지역을 회귀모형 추정을 위한 자료로 활용하였다. 실업률 추정을 위해 적합된 회귀모형은 다음과 같다.

$$\hat{Y} = 0.008 - 0.201X_1 + 0.680X_2 + 0.404X_3, \quad (3.2)$$

$$\text{Residual Mean Square} = 0.868 \times 10^{-4},$$

$$\text{추정치}의\ \text{표준오차} = 0.932 \times 10^{-2},$$

$$R^2 = 0.546$$

여기에서

Y = 5개월 간의 월별 CPS 실업률 추정값의 평균,

X_1 = 전체 실업자 대비 보험가입 실업자의 비율(%),

X_2 = 최종 공표된 연말 실업률(70-단계 추정값),

X_3 = “결혼여부-연령대별-성별-인종별”범주에서 계산된
실업률의 합성추정값

위 회귀모형의 타당성 검증을 위하여 센서스 자료를 이용한 회귀모형을 적합한 결과는 다음과 같이 주어지며, 실업률 추정에서 회귀모형 적용에 대한 타당성을 보여준다.

$$\hat{U} = 0.009 + 0.012X_1 + 0.586X_2 + 0.540X_3, \quad (3.3)$$

$$\text{추정치의 표준오차} = 0.213 \times 10^{-4},$$

$$R^2 = 0.859$$

여기에서 U 는 센서스 자료에서 계산된 실업률 추정값이다.

5개월 간의 월별 CPS 실업률 추정값들의 평균을 종속변수로 사용했을 때, 변동의 55%가 회귀모형으로 설명될 수 있고, 센서스의 실업률 추정값들을 종속변수로 사용한 경우에는 변동의 약 86%가 회귀모형으로 설명될 수 있다. CPS 표본 추정값들의 변동이 커질수록 설명변수들의 기여도가 낮아진다는 것을 확인할 수 있다.

다음은 독립변수들을 변화시켜 가면서 추가로 2종의 회귀모형을 추정해 본 결과이다. 먼저 독립변수 X_2 대신 벤치마킹이 이루어지기 전의 실업률에 대한 CPS 추정값 X_4 를 모형에 적합시킨 결과이다.

$$\hat{Y} = 0.012 - 0.252X_1 + 0.299X_3 + 0.078X_4, \quad (3.4)$$

$$\text{Residual Mean Square} = 0.833 \times 10^{-4},$$

$$\text{추정값의 표준오차} = 0.912 \times 10^{-2},$$

$$R^2 = 0.565$$

센서스 자료에 의해 위와 대응되는 회귀모형은 다음과 같다.

$$\hat{U}' = -0.006 + 0.005X_1 + 0.477X_3 + 0.564X_4, \quad (3.5)$$

$$\text{Residual Mean Square} = 0.228 \times 10^{-4},$$

$$\text{추정값의 표준오차} = 0.478 \times 10^{-2},$$

$$R^2 = 0.849$$

회귀모형 (3.4)는 총변동의 약 57%를 설명하며, 회귀모형 (3.2)보다는 어느 정도 개선된 효과가 나타났다.

두 번째 경우에는 최종 공표된 연말 실업률 X_2 대신 11개월 간의 CPS 월별 추정값들의 평균 X_5 와 “직업별-성별-인종별” 범주의 유사정보를 추정과정에 도입하여 구한 합성추정값 X_6 를 독립변수로 선택하였을 때 계산된 회귀모형의 적합 결과이다.

$$\hat{Y}'' = 0.009 - 0.210X_1 + 0.640X_5 + 0.444X_6, \quad (3.6)$$

$$\text{Residual Mean Square} = 0.883 \times 10^{-4},$$

$$\text{추정값의 표준오차} = 0.939 \times 10^{-2},$$

$$R^2 = 0.539$$

센서스 자료에 의한 적합 결과는 다음과 같이 주어진다.

$$\hat{U}''' = -0.008 - 0.011X_1 + 0.532X_5 + 0.617X_6, \quad (3.7)$$

$$\text{Residual Mean Square} = 0.194 \times 10^{-4},$$

$$\text{추정값의 표준오차} = 0.440 \times 10^{-2},$$

$$R^2 = 0.872$$

위의 적합결과를 살펴보면, 센서스 자료에 의한 추정값을 종속변수로 이용했을 때 회귀식은 약 85 ~ 87%의 설명력을 보이나, CPS자료 추정치를 사용했을 경우에 회귀식의 설명력은 약 54 ~ 56% 정도임을 확인할 수 있다.

다음 <표3.5>는 변수들 간의 상관관계를 나타낸 표이다. “70-단계” 추정값 X_2 는 센서스 추정값 U 와 CPS 추정값 Y 와 높은 상관관계를 보인다. 또한, 합성추정값 X_3 와 실직보험가입비율 추정값 X_1 과는 낮은 상관성을 나타내며, “70-단계” 추정값 X_2 와도 상관계수의 값이 낮다. 따라서 X_2 이외에 추가적인 독립변수로서 합성추정값 X_3 와 실직보험가입비율 추

정값 X_1 을 고려한다면 회귀식의 예측력은 증가할 것임을 알 수 있다.

<표3.5> 변수들의 가중상관계수

	X_1	X_2	X_3	X_4	X_5	X_6	Y	Z	U
X_1	1.000								
X_2	0.676	1.000							
X_3	0.285	0.512	1.000						
X_4	0.682	0.961	0.574	1.000					
X_5	0.666	0.995	0.477	0.959	1.000				
X_6	0.369	0.584	0.974	0.633	0.548	1.000			
Y	0.372	0.692	0.577	0.720	0.682	0.599	1.000		
Z	0.259	0.525	0.340	0.543	0.521	0.339	0.700	1.000	
U	0.554	0.851	0.756	0.868	0.815	0.810	0.741	0.512	1.000

* $Z=1970$ 년 5월 CPS 실업률 추정값

센서스 중간 해에는 회귀모형의 종속변수는 CPS 1차 추출단위에 대한 추정값이 되며, 독립변수로 센서스에서 추정된 실업률 U 를 선택할 수 있으며, 적합된 회귀식은 다음과 같다.

$$\hat{Y} = 0.010 + 0.450U + 0.326X_4 + 0.089X_6, \quad (3.8)$$

$$\text{추정값의 표준오차} = 0.914 \times 10^{-4},$$

$$R^2 = 0.563$$

만약 종속변수로 1970년 5월 CPS 실업률 추정값 Z 를 사용한다면 적합된 회귀모형은 다음과 같다.

$$\hat{Z} = 0.019 + 0.422U + 0.430X_4 - 0.246X_6, \quad (3.9)$$

$$\text{Residual Mean Square} = 2.040 \times 10^{-4},$$

$$\text{추정값의 표준오차} = 1.428 \times 10^{-2},$$

$$R^2 = 0.291$$

회귀적합모형 (3.8)은 독립변수로 센서스 추정값을 이용한 결과이다. 회귀모형은 약 56%의 설명력을 가지며, 또한 Y를 종속변수로 갖는 다른 회귀적합모형과 비교했을 때 결코 떨어지지 않는 설명력을 갖는다. 1970년 5월 CPS 실업 추정치를 종속변수로 선택한 회귀적합모형 (3.9)은 약 29%의 설명력을 갖는다. 이러한 낮은 설명력은 종속변수의 큰 분산(변동)에 기인한다고 볼 수 있다.

노동력의 규모가 서로 상이한 지역들에 대한 잔차분석 결과가 <표 3.6>에 주어졌다. 독립변수는 각각 (X_1, X_3, X_4) , (X_1, X_5, X_6) 이고, 종속변수로써 각각 1970년 센서스 추정값, 5개월간의 월별 CPS 추정값, 1970년 5월의 CPS 추정값을 이용한 회귀모형에 대한 잔차의 평균과 잔차의 표준오차이다.

<표3.6> 잔차의 평균과 표준오차 (단위:%점)

1970센서스 노동력크기	SMSA 의 수	센서스실업률		CPS 5개월 평균실업률		CPS 5월 실업률(1970년)	
		잔차 평균	잔차 표준오차	잔차 평균	잔차 표준오차	잔차 평균	잔차 표준오차
(X_1, X_3, X_4)							
1,000,000 이상	9	-0.18	0.31	-0.01	0.39	-0.00	0.54
500,000-999,999	17	-0.02	0.54	-0.06	0.66	0.24	0.88
250,000-499,999	21	0.15	0.36	0.15	1.17	-0.21	1.23
100,000-249,999	51	0.14	0.66	0.03	1.55	0.10	2.77
100,000 미만	24	0.37	0.78	0.41	1.83	-0.46	2.83
(X_1, X_5, X_6)							
1,000,000 이상	9	-0.07	0.34	0.07	0.38	0.06	0.55
500,000-999,999	17	-0.10	0.49	-0.16	0.75	0.10	0.90
250,000-499,999	21	0.06	0.39	0.06	1.23	-0.28	1.37
100,000-249,999	51	0.10	0.61	0.03	1.57	0.14	2.80
100,000 미만	24	0.11	0.71	0.19	1.75	-0.63	2.86

전반적으로 노동력의 규모가 작은 지역들보다는 노동력의 규모가 큰

지역들에서 추정값의 효율이 좋다. 잔차의 표준오차는 노동력의 규모가 작을수록 증가하는 경향을 보인다. 센서스 추정값을 종속변수로 취한 회귀모형에서 노동력의 규모가 100,000이하인 지역을 살펴보면 잔차의 표준 오차는 노동력의 규모가 백만 이상인 지역보다 2배 이상 큰 표준오차를 갖는다. 5개월 간의 월별 CPS 평균에 대한 추정값을 이용한 회귀모형이 1970년 5월 CPS 추정값을 이용한 회귀모형 보다 잔차의 표준오차가 훨씬 작게 나타난다.

1970년 센서스 실업률과 회귀모형 추정값 간의 차이에 대한 분포가 다음 <표3.7>에 주어졌다. 독립변수는 (X_1, X_5, X_6) 를 이용하였다. CPS 추정값을 이용한 경우보다 센서스 실업률 추정값을 이용했을 경우가 훨씬 대칭적인 분포를 보인다.

<표3.7> 센서스실업률과 회귀추정값의 차이에 대한 분포
(122개의 SMSA 이용)

차이의 범주 (%점)	센서스 실업률		5개월 간의 CPS 평균 실업률		1970년 5월의 CPS 실업률	
	SMSA의 갯수	백분율 (%)	SMSA의 갯수	백분율 (%)	SMSA의 갯수	백분율 (%)
3.00 이상	0	0	0	0	0	0
2.00-3.00	1	0.8	1	0.8	1	0.8
1.50-2.00	0	0	0	0	0	0
1.00-1.50	6	4.9	0	0	2	1.6
0.50-1.00	14	11.5	4	3.3	5	4.1
0.25-0.50	15	12.3	5	4.1	8	6.6
0.10-0.25	19	15.6	1	0.8	3	2.5
-0.10-0.10	23	18.9	9	7.4	6	4.9
-0.25 ~ -0.10	10	8.2	11	9.0	8	6.6
-0.50 ~ -0.25	15	12.3	15	12.3	16	13.1
-1.00 ~ -0.50	16	13.1	47	38.5	43	35.2
-1.50 ~ -1.00	3	2.5	25	20.5	23	18.9
-2.00 ~ -1.50	0	0	2	1.6	5	4.1
-3.00 ~ -2.00	0	0	2	1.6	2	1.6
-3.00 이하	0	0	0	0	0	0

-1.0에서 1.0까지의 구간을 살펴보면, 센서스 실업률을 종속변수로 취한 회귀모형에서는 약 91.9%의 SMSA지역들이 이 구간 내에 포함되며, 5개월 간의 CPS 추정값의 평균을 종속변수로 취한 회귀모형에서는 약 75.3%, 1970년 5월 CPS 추정값을 종속변수로 취한 회귀모형에서는 약 73.0%의 SMSA지역들이 포함되어 있다. -0.50 이하의 범주에서는 5개월 간의 CPS 실업률 추정값의 평균을 종속변수로 취한 회귀모형이 약 62.2%, 1970년 5월 CPS 실업률 추정값을 종속변수로 취한 회귀모형이 약 59.8%의 SMSA지역들을 포함하고 있으며, 이 범주에서는 센서스 실업률을 종속변수로 취한 회귀모형보다는 CPS 추정값을 종속변수로 취한 회귀모형이 더 많은 SMSA지역들을 포함하고 있다.

5개월 간의 월별 CPS 추정값의 평균을 종속변수로 취한 회귀모형에서

$$MSE = E\left\{\frac{(Y_0 - \hat{Y})(Y_0 - \hat{Y})}{n}\right\} = \frac{(n - 2p - 2)\sigma_w^2}{n}$$

을 이용하여 계산된 MSE 추정결과는

독립변수	MSE
X_1, X_2, X_3	0.405×10^{-4}
X_1, X_3, X_4	0.369×10^{-4}
X_1, X_5, X_6	0.419×10^{-4}

와 같이 주어진다. 여기에서 “ Y_0 =관측값 Y ”, “ n =관측값의 개수”, “ p =독립변수의 개수”, “ σ_w^2 =PSU 내의 오차”를 나타낸다.

상기한 자료들은 다음과 같은 내용을 시사한다. 미 노동통계국에서 연말 공표하는 실업률 추정값은 CPS의 1차 추출단위 추정값을 독립변수로 취한 회귀모형을 이용하여 개선될 수 있음을 시사한다. 추가적인 독립변수들로 70-단계 추정값 외에 UI자료, 센서스자료, CPS 자료를 이용한 합성추정값, 센서스 중간년도의 표본 자료 추정값 등을 이용할 수 있다. 회귀모형에 의한 소지역 실업 통계의 추정은 적절한 독립변수와 종속변수의 선택이 중요한 과제이지만, 센서스와 CPS자료를 이용한 전반적인

검토 및 실업에 관련된 사회경제적 요인에 대한 파악도 중요한 과제이다.

(2) 시계열 회귀모형

시계열 회귀모형은 현재 미국에서 표본크기가 충분하지 못한 주 단위 및 주 내의 부차관심영역들의 실업통계를 작성하는데 적용하고 있는 방법으로 표본조사 추정량을 개선하기 위해 모집단의 값을 확률과정으로 생각하여 시계열 분석의 시그널 추출 기법을 적용한 것이다.

1989년 1월 노동통계국(BLS)에서는 39개 주와 콜롬비아 구의 총 40개 지역에 대해 월별 취업 및 실업자 추정을 위한 새로운 방법을 소개하였다. 이 방법은 CPS의 월별 표본자료에 시계열 모형을 적합시키는 방법이다.

모집단의 특성을 추정하기 위한 직접적인 방법은 표본 설계에 근거하여 대규모 표본조사를 시행하는 것이다. 대부분 추정값들은 대 영역에 대해서는 신뢰할 만한 수준이나 대 영역 내의 부차관심영역들에 대해서는 반드시 그러하지는 않다. 주기적인 조사의 경우 소규모의 주 지역 및 주 내의 부차관심영역들에 대해서 시계열 기법이 관심을 받고 있으며, CPS 추정값은 이러한 기법을 적용시키기에 적합한 자료이다. 매달 약 50,000 표본 가구가 조사되어 모집단의 노동력 상태를 추정하는데, 전국 수준의 추정값 또는 인구수가 많은 11개 주에 대한 월별 추정값은 비교적 신뢰할 만 하다. 인구수가 적은 나머지 40개 지역에 대해서는 월별 CPS 추정값을 그대로 사용하는 것은 바람직하지 못하다.

새로운 추정은 CPS 추정값을 시간에 따라 확률적으로 변화하는 시계열로 간주하여 시그널 성분과 노이즈 성분의 합으로 표현되는 시계열 모형 추정 방법에 근거한다. 월별 CPS 추정값은 시계열 모형에서 실업보험(UI: Unemployment Insurance)자료와 경상고용통계(CES: Current Employment Statistics)자료를 결합하여 계산된다. 즉, 이전에 추정된 방법과는 달리 좀더 체계적인 방법으로써 이용 가능한 보조 자료와 과거와 현재의 표본 자료를 모두 이용하여 표본크기가 작을 때에서 발생하는 CPS 추정값의 추

정오차를 줄이고자 하는 것이 기본적인 생각이다.

비 관측된 모집단 값의 동적인 변화와 표본오차의 자기공분산을 나타내는 모형이 주어진다면, KF(Kalman Filter)는 참값을 추정하기 위한 방법으로 적용될 수 있으며, 다음과 같은 몇가지 유용성을 갖는다. 첫째, KF는 시그널 성분과 노이즈 성분으로 표현된 모형들에 대한 다양한 추정을 허용한다. 둘째, KF의 반복적인 추정 방법은 주 및 주 내의 부차관심 영역들에 대한 월별 노동력 관련 추정값을 산출하는 데에 있어서 매우 효율이 좋은 알고리즘을 제공한다. 셋째, KF는 동적인 모형들에 대해서 미지인 모수들을 효율적으로 추정하기 위한 매우 유용한 도구이다.

① CPS 자료를 모형화하기 위한 시계열 방법

CPS 노동력 추정값 $y(t)$ 는 시그널 성분 $\theta(t)$ 와 노이즈 성분 $e(t)$ 의 합인

$$y(t) = \theta(t) + e(t)$$

로 주어진다. 시그널 성분(노동력의 참값)은 시간에 따라 변화하는 $\mu_x(t)$ 와 오차항 $u(t)$ 로 표현할 수 있다.

$$\begin{aligned} \theta(t) &= \mu_x(t) + u(t) \\ &= X(t)\beta(t) + u(t). \end{aligned} \tag{3.10}$$

여기에서 $X(t)$ 는 $1 \times k$ 벡터로써 관측된 값이며, $\beta(t)$ 는 $k \times 1$ 확률계수 벡터이다. 회귀계수 $\beta(t)$ 는 1차 자기상관회귀 과정에 따라 확률적으로 변화하는 형태를 수식으로 표현하면

$$\beta(t) = T_\beta \beta(t-1) + v_\beta(t) \tag{3.11}$$

로 주어진다. 여기에서 T_β 는 $k \times k$ 행렬로 고정모수들의 행렬을 나타내며, $v_\beta(t)$ 는 $k \times 1$ 백색잡음오차 벡터를 나타낸다. 식(3.10)에서 오차항 $u(t)$ 는 자기상관성을 고려하여 $u(t) \sim ARMA(p_u, q_u)$ 로 가정되며

$$u(t) = \phi_u(L)\psi_u(L)^{-1}v_u(t) \tag{3.12}$$

와 같이 표현된다.

여기에서

$v_u(t) = u(t)$ 에 대한 백색잡음오차,

$$\psi_u(L) = 1 - \sum_{i=1}^{p_u} \psi_u(i)L^i, \quad (u(t)\text{에 대한 자기회귀연산자})$$

$$\phi_u(L) = 1 + \sum_{i=1}^{q_u} \phi_u(i)L^i, \quad (u(t)\text{에 대한 MA연산자})$$

이고 L 은 시차연산자로서 $L^i y(t) = y(t-i)$ 를 만족한다. 랜덤오차 $v_\beta(t)$ 와 $v_u(t)$ 는 평균이 0이고 서로 독립이라고 가정한다. 즉,

$$\begin{pmatrix} v_\beta(t) \\ v_u(t) \end{pmatrix} \sim ID \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & \sigma_{v_u} \end{pmatrix} \right) \quad (3.13)$$

단, $Q = Cov(v_\beta(t)) = Diag(\sigma_{\beta\beta}, \dots, \sigma_{\beta\beta})$.

노이즈 성분은 전체 모집단에서 표본 일부를 추출하는 과정에서 발생하는 표본오차로 취급된다. CPS는 모집단으로부터 추출된 다 단계 확률 표본이다. 일차적으로 모집단으로부터 일차추출단위(PSU)들이 뽑히고, 다음으로 PSU 내에서 조사가구 단위들이 추출된다.

$y(t)$ 가 노동력 특성을 나타내는 $\theta(t)$ 에 대한 CPS 추정치로 주어지면, 표본오차 $e(t) = y(t) - \theta(t)$ 의 분산과 공분산 함수는 다음과 같다.

$$\sigma_{e(t)e(t)} = D_y S_y^2 \quad (3.14)$$

$$\gamma_{ts} = Cov(e(t), e(s))$$

여기에서

D_y = 단순임의추출표본 추정치의 분산에 대한 CPS 추정치의 분산의 비(Design Effect),

$$S_y^2 = \frac{N(t)}{n(t)} \theta(t) (1 - P(t)),$$

$N(t)$ = 모집단의 크기,

$n(t)$ = 표본의 크기,

$$P(t) = \frac{\theta(t)}{N(t)}.$$

위의 식에서 설명된 것과 같이 CPS 표본오차는 이분산적 구조와 자기상관적 구조를 동시에 갖는다. 식(3.14)는 표본 재설계에서 설계효과 D_y , 표본구간 $N(t)/n(t)$, 참값 $\theta(t)$ 와 $P(t)$ 의 변화와 같은 이분산성의 세가지 주요한 요인이 반영된 식이다. 부연하여 설명하면, CPS는 10년 간의 센서스 자료를 사용하여 매 십년 마다 표본 추출 구조 및 추정 절차를 갱신하기 위하여 재 설계된다. 표본조사의 재 설계보다는 오히려 주 단위에 대한 표본 크기의 조정이 자주 있었고, 이것은 주 수준의 분산에 중요한 영향을 끼치게 되었다. 일종의 고정된 설계와 고정된 표본 크기일지라도 오차 분산은 참 노동력 크기의 함수이므로 변하게 된다. 노동력은 매우 순환적이면서 동시에 계절적이므로 분산도 이와 비슷한 형태를 가진다는 것을 예측해야만 한다.

$e(t)$ 의 자기공분산 구조는 다음의 세 가지 사실에 기인한다. 첫째, 월별 표본은 8개의 서로 독립적인 연동교체그룹들인 부 표본들로 구성되어 있다. 연동교체그룹들은 4개월 간 연속 조사되고, 8개월 동안 조사되지 않다가 다시 4개월 간 연속 조사 후 표본에서 영구히 제거된다. 이 과정에서 동일한 조사가구 단위들이 시간에 따라 중복되어 조사되므로 당연히 상관성이 존재한다. 4-8-4 연동교체표본 체계는 월별 추정값의 신뢰성을 높이기 위해 15개월 주기 동안 첫 달은 전월에 조사된 표본의 약 75%를 중복하여 월 표본으로 사용하고, 나머지는 교체표본 추출방식에 의해 새로운 표본을 조사한다. 2개월 간의 중복율은 약 50%정도이다. 둘째, 연동교체표본 체계의 사용은 표본들에 대한 주기적 선택을 요구한다. 한 집락의 조사가구 단위들이 연동교체표본 그룹으로부터 영구히 제외될 때, 근처에 있는 단위들로 대체된다. 따라서 새로운 단위들도 대체 단위들과 비슷한 특성을 가질 것이므로 같은 연동교체표본 그룹에서 다른 가구들과도 높은 상관성을 갖게 된다. 셋째, 표본오차는 복합추정량에 의해 영향을 받게 된다.

CPS 추정량의 이분산적이며 자기상관적인 구조는 $e(t)$ 를

$$e(t) = \gamma(t)e^*(t)$$

와 같은 승법적 구조로 모형화 함으로써 설명될 수 있다(Bell and Hillmer(1989)). 여기에서 $e^*(t)$ 는 ARMA과정을 따르며 상수인 분산을 갖는다.

$$e^*(t) = \phi_e(L)\psi_e(L)^{-1}v_e(t) \quad (3.15)$$

여기에서

$$v_e(t) \sim NID(0, \sigma_{v_e}),$$

$$\sigma_{e^*(t)e^*(t)} = \sigma_{v_e} \sum_{k=0}^{\infty} g_k^2$$

이며, 가중치 $\{g_k\}$ 는 생성함수 $g(L) = \phi_e(L)\psi_e(L)^{-1}$ 로 계산된다. $e(t)$ 의

이분산성 성분 $\gamma(t) = \sqrt{\frac{\sigma_{e(t)e(t)}}{\sigma_{e^*(t)e^*(t)}}}$ 는 분산비의 제곱근이다. $e(t)$ 의 자기상

관구조는 표본 설계에 의해 영향을 받게 된다. 예를 들면, 표본 재설계에서 복합추정량에서의 가중치들이 모두 교체되는 것과 같은 경우이다.

② 상태공간형식(State Space Form)과 KF 알고리즘

노동력 모형의 시그널 성분과 노이즈 성분을 상태공간형식에 삽입한다. 먼저 추정에는 적절치 못할지라도 융통성을 보여 주기에는 유용할 것 같은 매우 일반적인 형식을 설명하고 실질적인 적용에 있어서는 고려되어야 할 것 같은 제한사항들에 대해서 설명한다. 상태공간 작성에서는 비관측된 시그널과 노이즈가 상태 변수들이고, 이들의 시간에 따른 전개는 변이함수들의 집합에 의해 설명된다. 관측점 함수는 상태 변수들을 관측 표본시계열들로 변환시킨다. 상태공간 체계에서 변이함수들은 1계의 VAR 형태를 취한다.

비 관측변수는 $\beta(t)$, $u(t)$, $e^*(t)$ 이다. 계수벡터인 $\beta(t)$ 는 (3.11)식과

같이 이미 적절한 형식이 있다. $u(t)$ 와 $e^*(t)$ 는 (3.12)식과 (3.15)식에서와 같이 $ARMA$ 과정으로 명시되나, 각각 1계 VAR 형태인 $S_u(t)$ 와 $S_e(t)$ 벡터들로 변환된다. 임의의 $ARMA(p, q)$ 과정은 일종의 $r \times 1$ 1계 VAR 형태로 변환될 수 있다는 것이 기본적인 규칙이며, 여기에서 $r = \max(p, q + 1)$ 이다(Harvey(1981)). 변이함수들은 다음과 같이 주어지며, 여기에서 $S(t)$ 는 $\beta(t)$, $S_u(t)$ 와 $S_e(t)$ 로 구성되는 상태벡터이다.

$$S(t) = T(t)S(t-1) + \Gamma(t)v(t),$$

$$(m \times 1)(m \times m) \quad (m \times l)(l \times 1)$$

$$\begin{pmatrix} \beta(t) \\ S_u(t) \\ S_e(t) \end{pmatrix} = \begin{pmatrix} T_\beta & 0 & 0 \\ 0 & T_u & 0 \\ 0 & 0 & T_e(t) \end{pmatrix} \begin{pmatrix} \beta(t-1) \\ S_u(t-1) \\ S_e(t-1) \end{pmatrix} + \begin{pmatrix} I_k & 0 & 0 \\ 0 & \Gamma_u & 0 \\ 0 & 0 & \Gamma_e(t) \end{pmatrix} \begin{pmatrix} v_\beta(t) \\ v_u(t) \\ v_e(t) \end{pmatrix},$$

$$E(v(t)v(t)') = \text{block diagonal}(Q, \sigma_{u_u}, \sigma_{v_{e(t)}v_{e(t)}}).$$

여기에서

$$T_u = \begin{pmatrix} \psi_u(1) & : & I_{r_u-1} \\ \psi_u(2) & : & \\ \vdots & \vdots & \\ \psi_u(r_u) & : & 0 \end{pmatrix}, T_e(t) = \begin{pmatrix} \psi_e(1) & : & I_{r_e-1} \\ \psi_e(2) & : & \\ \vdots & \vdots & \\ \psi_e(r_e) & : & 0 \end{pmatrix}$$

$$\Gamma_u = \begin{pmatrix} 1 \\ \phi_u(1) \\ \phi_u(2) \\ \vdots \\ \phi_u(r_u-1) \end{pmatrix}, \Gamma_e = \begin{pmatrix} 1 \\ \phi_e(1) \\ \phi_e(2) \\ \vdots \\ \phi_e(r_e-1) \end{pmatrix},$$

$$m = k + r_u + r_e,$$

k = 회귀변수의 수,

$$l = k + 2,$$

$$r_u = \max(p_u, q_u + 1), \quad p_u \text{와 } q_u \text{는 } u(t) \text{의 } ARMA \text{모수,}$$

$r_e = \max(p_e, q_e + 1)$, p_e 와 q_e 는 $e(t)$ 의 ARMA모수.

관측점 함수는 벡터 $H(t)$ 를 선택하여, 시그널 성분과 노이즈 성분을 만들기 위한 상태변수들의 일차결합을 취한다.

$$Y(t) = H(t)S(t) = \theta(t) + e(t), \quad (3.16)$$

여기에서

$$H(t) = (X(t) | 1 | 0_{r_e-1} | \gamma(t) | 0_{m-k-r_e-2}),$$

$$\theta(t) = H_\theta(t)S(t),$$

$$e(t) = H_e(t)S(t),$$

$$H_\theta(t) = (X(t) | 1 | 0_{m-k-1}),$$

$$H_e(t) = (0_{k+r_e} | \gamma(t) | 0_{r_e-1}).$$

비 관측된 시그널과 노이즈 성분들의 상태공간형태가 주어진다면, KF는 시그널과 노이즈를 추정하기 위한 방법을 제공한다. 이러한 알고리즘을 설명하기 위해 시간 $t-j$ 까지 관측된 자료에 대한 $S(t)$ 의 조건부 평균 및 공분산 행렬을 다음과 같이 표현하자.

$$S(t|t-j) = E(S(t) | Y_{t-j}, \dots, Y_1),$$

$$P(t|t-j) = E\{(S(t) - S(t|t-j))(S(t) - S(t|t-j))^t | Y_{t-j}, \dots, Y_1\}.$$

또한 바로 전까지의 값들이 주어진 경우의 표본 추정값 $Y(t)$ 의 예측값을

$$Y(t|t-1) = H(t)S(t|t-1),$$

$Y(t)$ 의 분산을

$$\begin{aligned} E(Y(t) - Y(t|t-1))^2 &= H(t)P(t|t-1)H(t)^t \\ &= f(t|t-1) \end{aligned}$$

로 표현하자.

t 번째 관측점까지의(t 번째 관측점은 제외) 자료에 근거한 $S(t)$ 의 추정값이 주어진다면, 최근의 자료를 고려한 $S(t)$ 의 추정량은 $S(t|t-1)$ 과 최근의 표본 추정값 $Y(t)$ 의 가중평균으로

$$S(t|t) = (I - K(t)H(t))S(t|t-1) + K(t)Y(t)$$

$$= S(t|t-1) + K(t)(Y(t) - Y(t|t-1))$$

와 같이 표현되며, 공분산 행렬은

$$P(t|t) = (I - K(t)H(t)^t)P(t|t-1)$$

로 주어지며,

$$S(t|t-1) = TS(t-1|t-1),$$

$$P(t|t-1) = TP(t-1|t-1)T^t + \Gamma E(v(t)v(t)^t)\Gamma^t$$

로부터 반복적으로 추정된다. 여기에서 이득벡터 $K(t)$ (gain of filter)는

$$K(t) = \frac{P(t|t-1)H(t)^t}{f(t|t-1)}$$

로 표현되며, $K(t)$ 의 원소들은 $P(t|t)$ 의 대각원소들의 합을 최소화하여 결정한다(Gelb(1974)).

KF 방정식을 이용하여 시간 t 에서 관측된 표본 추정값은 시그널 성분과 노이즈 성분으로 분해된다.

$$Y(t) = \theta(t|t) + e(t|t).$$

여기에서

$$\theta(t|t) = \theta(t|t-1) + h_\theta(t)\tilde{Y}(t),$$

$$e(t|t) = e(t|t-1) + (1 - h_\theta(t))\tilde{Y}(t),$$

$$\tilde{Y}(t) = Y(t) - Y(t|t-1),$$

$$h_\theta(t) = H_\theta(t)K(t)$$

$$= \frac{\text{Var}\left(\frac{\theta(t)}{\theta(t-1)}\right) + H_\theta(t)TP(t-1|t-1)T^tH(t)}{f(t|t-1)},$$

$$1 - h_\theta(t) = H_e(t)K(t)$$

$$= \frac{\sigma_{e(t)e(t)} + H_e(t)TP(t-1|t-1)T^tH(t)}{f(t|t-1)},$$

$$\text{Var}\left(\frac{\theta(t)}{\theta(t|t-1)}\right) = \sum_{i=1}^k X_i(t)^2 \sigma_{\beta_i} + \sigma_{u_u},$$

$$\sigma_{e(t)e(t)} = \gamma(t)^2 \sigma_{v,t} .$$

가중치 $h_\theta(t)$ 는 예측오차 $\tilde{Y}(t)$ 를 시그널과 노이즈 성분으로 분해하며, 이것은 KF 가 시그널 성분의 최소평균제곱오차 추정값을 만들기 위해 시계열 추정값 $\theta(t|t-1)$ 과 최근의 표본 추정값 $Y(t)$ 를 결합하는 구체적인 방법을 설명한다. $\theta(t|t-1)$ 의 $Y(t)$ 에 대한 보정 양은 이분산성의 표본오차 분산 $\sigma_{e(t)e(t)}$ 와 상대적으로 비교하여 시계열 분산성분 $Var\left(\frac{\theta(t)}{\theta(t-1)}\right)$ 의 크기의 함수로 표현된다. $\sigma_{e(t)e(t)}$ 의 값이 크면 $h_\theta(t)$ 의 값이 작게 되며, 따라서 $\theta(t|t)$ 를 이끌어 내는데 있어서 시계열 예측값 $\theta(t|t-1)$ 에 대한 작은 양의 보정만이 일어난다. 역으로 만약 표본 분산이 작다면 $\theta(t|t)$ 는 최근의 표본 추정값 $Y(t)$ 와는 아주 다른 값이 될 것이다.

KF 는 반복적인 방법으로 상태벡터 $S(t)$ 의 최소평균제곱오차를 제공하며, 또한 KF 는 새로운 자료가 각 주기에서 이용될 수 있는 실시간 상황에 적합하다. 그러나 자료가 시간 t 이후에 이용될 수 있을 경우에는 추정값 $S(t|t)$ 는 이러한 새로운 정보를 반영시키지는 못한다. 왜냐하면 KF 는 시간 t 의 앞쪽으로부터 이동하기 때문이다. 이 전 주기의 추정값들에 대한 부분적인 최적화는 평활을 통해 쉽게 조정될 수 있다.

평활의 방법은 두 가지 형태의 KF 추정량을 결합하는 방법이다. 첫 번째 형태의 KF 추정량은 이 전에 설명했던 것과 같이 전방 Filter 추정값으로써 시간 t 에서 모든 과거 표본자료와 현재 표본자료에 의해 추정되는 $S(t|t)$ 이다. 두 번째 형태의 KF 추정량은 후방 Filter 추정치으로써, 표본 주기의 끝에서 출발하여(가령 $t=n$ 에서 출발) 처음까지 진행하며, 미래의 자료에만 근거하여, 각 시간 t 에서의 예측값들을 이끌어 내는데, 이러한 추정량을 $S(t|t+1)$ 로 나타낸다. 이 때 최적인 평활 추정량은 두

추정량의 평균제곱오차 값들의 비율을 결합하여 작성되며 구체적인 형태는 다음과 같이 주어진다.

$$S(t|n) = P(t|n) \left\{ \frac{S(t|t)}{P(t|t)} + \frac{S(t|t+1)}{P(t|t+1)} \right\}.$$

여기에서

$$P(t|n) = 1 / \left\{ \frac{1}{P(t|t)} + \frac{1}{P(t|t+1)} \right\}.$$

한편, $S(t|n)$ 의 공분산 표현식으로부터

$$\frac{1}{P(t|n)} = \frac{1}{P(t|t)} + \frac{1}{P(t|t+1)}$$

을 얻을 수 있고 $P(t|n) - P(t|t)$ 는 음 반정치(negative semidefinite)임을 알 수 있다. 따라서 $S(t)$ 의 평활 추정량은 전방 Filter추정량보다는 훨씬 좋은 추정량이 되며, 이러한 이유 때문에 노동력 추정값들은 평활 알고리즘을 이용하여 만들어진다.

③ 보충 설명

상태공간형식은 시그널 성분을 세분화하는데 상당한 유연성을 허용하며, 특별한 경우들로써 표본조사 방법에 근거한 다음의 두 가지 형태의 모형을 포함한다.

만약 $Q = Cov(v_\beta(t)) = 0$ 이고 $e(t)$ 와 $u(t)$ 가 백색잡음오차 변수이면, 시스템은 Ericksen(1974)의 표본회귀모형이 된다. 이 경우 시그널 적출 문제는 관측된 표본자료에 대해 가중 최소제곱방정식을 적합시켜 해결된다. $\beta(t) = 0$, $Q = 0$ 일 경우 Wiener-Kolmogorov의 시그널 적출 이론에 근거한 모형을 얻는다. 회귀식의 평균은 없어지고 시그널은 공분산 정상과정이 된다. 추가적으로 $e(t)$ 과정의 분산과 ARMA 모수들이 상수인 경우에는 $e(t)$ 가 공분산 정상과정이 된다.

Scott와 Smith(1974)는 공분산이 추정되어야 하는 조사 자료에 대해 전통적인 시그널 적출 방법을 채택하였다. Bell과 Hillmer(1987a)는 시그

널 과정에서 비 정상성을 다루는 방법들을 토의하였다. 앞서 소개되었던 내용에서는 시그널의 비 정상성 문제를 회귀변수들과 회귀계수들의 확률적인 변화로 다루었다. 회귀계수들의 움직임을 통제하는 (3.11)식의 추이 방정식은 다양한 형태를 수용할 수 있다(Los(1985)).

CPS 자료를 이용한 연구는 많이 진행되고 있지는 않다. Hausman과 Watson(1985)은 CPS 4-8-4 연동표본교체와 합성추정절차를 통합시킨 전 지역의 십대의 실업률 시계열에 대한 오차과정의 $ARMA(1, 15)$ 모형을 개발했다. Bell과 Hillmer(1987b)는 Train(1978)등에 의해 추정된 design에 근거한 자기공분산의 근사으로써 $ARMA(1, 1)$ 모형을 개발했다.

CPS와 같은 복잡한 조사에 대해서 공분산 추정값을 산출하는 것은 대규모의 자료가 사용되어 비용이 많이 든다. 최근에는 필요로 하는 모든 자료를 이용하지 않고 추정값을 산출하는 방법이 모색되고 있다. 실제 기반 표본오차 공분산 계산의 어려움 때문에 방정식에서의 오차들의 효과와 표본오차의 효과를 추정하지 않고 회귀방정식을 적합시키기도 한다. 만약 두 성분 오차들이 $ARMA$ 과정이면, 이 때 합도 $ARMA$ 과정이 된다 (Granger and Morris(1975)). 즉,

$$u(t) \sim ARMA(p_u, q_u), e(t) \sim ARMA(p_e, q_e) \text{ 이면}$$

$$w(t) = u(t) + e(t) \sim ARMA(p, q).$$

여기에서 $p \leq p_u + p_e$, $q \leq \max(p_u + q_e, p_u + q_e)$ 를 만족한다. KF 는 회귀성분과 총 오차를 적출하기 위해 사용될 수 있다.

④ 실업률 추정에 적용

CPS 자료 외의 미 연방-주의 실업보험 체계로부터 생산된 실업보험(UI) 자료와 비농인구에 대한 경상고용통계조사(CES) 자료가 있다. 이러한 자료들은 1960년대 초 이래로 주의 특정 추정치를 생산하기 위하여 미 노동통계국에서 작성한 소책자 추정방법에서 이용되었던 자료들이다. 이러한 자료들을 이용하여 순환적이며 계절적인 노동력의 움직임을 통제하

고 표본오차의 영향을 줄이기 위한 시계열 추정방법이 적용되었다. 이러한 방법은 구성된 변수들의 오차 문제를 다루기 때문에, 즉 계수들보다는 오히려 종속변수의 참값을 추정하는 데에 초점이 맞추어져 있기 때문에 기존의 방법과는 다르다.

보험가입 실직자 수에 대한 월별 주(state) 단위 자료는 CPS 자료와는 독립적으로 얻을 수 있는 실직자에 대한 최신정보이다. 이러한 자료가 실업률 모형 개발을 위한 출발점이 된다. UI 자료는 UI 혜택을 신청하고 있는 노동자들의 수를 완전히 집계한다. UI 보상은 해고되어 주의 특정한 재정상의 적격 기준을 충족시킨 노동자들에게만 혜택이 돌아간다. 대조적으로 CPS에서 이용되는 실직의 개념은 조사기간 동안 직업을 갖고 있지 않은 모든 사람들, 해고되어 직업을 찾고 있는 사람 또는 일시 해고되어 대기발령 중인 사람들을 모두 포함한다. UI 자료에 집계되지 않은 실직자 범주를 살펴보면 <표3.8>와 같다.

<표3.8> UI자료에 집계되지 않은 실직자 범주

- | |
|---|
| <ol style="list-style-type: none"> 1. 다음의 범주에 해당되는 실직자 <ol style="list-style-type: none"> a. Exhaustees : 수혜 권리를 다 써버린 노동자들 b. 재정적 부적격자 : 주의 적격 요구기준을 충족시키지 못하는 우선 고용자 또는 소득을 갖고 있는 노동자들 c. 유예된 비 신청자 : 실직시기의 처음에 혜택을 신청하지 않은 적격한 노동자들 2. 노동력 신규 유입자 : 최근의 실직기간 전에 노동인구에 편입되지 않았던 노동자들 3. 이직자 : 이 전의 직업을 그만두고 다른 직업을 찾고 있는 노동자 |
|---|

UI 자료에 집계되지 않은 실업자 중 신규 유입자가 차지하는 비율이 가장 크며, 다니는 직장을 그만두고 다른 직업을 찾는 이직 희망자들은 적어도 그 기간만큼은 보험 상의 혜택을 받지 못하는 못한다. 만약 UI 보상을 받지 못하는 실직자의 상대적인 규모가 시간에 따라 안정적이라면, 보험

지불요구율은 전체 실업률 추정에 대응으로 사용될 수 있다. 실제로 노동 시장에서 특히 실직자와 신규 유입자 간의 실업률 분포는 순환적이며 계절적인 특성변화를 나타낸다.

먼저 실업률 분포에 있어서 계절적 변화를 고려해 보자. 가장 중요한 현상은 실직자와 유입자는 매우 다른 계절적 형태를 띤다는 점이다. 청년과 여성이 유입자의 대부분을 차지하고, 청년층의 실업은 휘발성의 계절적 형태를 띠는데, 이는 학기가 끝나면서 신규 노동력이 유입되고 학기 시작과 함께 빠져나가는 청년층의 노동력의 변화가 반영되었기 때문이다. 이와는 대조적으로 성인 남성의 실직에 있어서 가장 일반적인 원인은 자동차 산업과 건설 산업과 같이 산업의 연간 생산 순환주기에 영향을 받아 계절적인 해고와 재 고용이 발생한다. <표3.9>는 CPS 신규유입 및 실직률의 계절적 형태(40개 주에 대한 평균값)와 전체 비율에 대한 이러한 항들의 순수효과 간의 차이를 나타내며, 또한 UI 가입자에 대한 계절적인 형태를 보여준다.

<표3.9> 40개 주에 대한 실업의 계절적인 요인들에 대한 평균 (1979-1985년)

월	CPS 실업률				UI 지불요구율
	전체	신규유입	실직	이직	
1	110.6	99.1	120.9	104.3	131.2
2	110.9	98.1	126.1	100.7	133.8
3	105.3	96.2	115.7	95.5	120.8
4	97.0	90.7	103.5	91.3	104.0
5	93.2	96.4	91.0	92.8	91.0
6	106.0	127.2	89.6	93.6	85.8
7	98.7	106.0	90.8	101.0	95.9
8	98.5	102.3	92.1	110.7	90.2
9	95.5	102.7	84.7	112.8	77.6
10	93.6	98.3	88.0	107.0	79.3
11	94.9	94.5	93.5	101.3	87.7
12	95.9	88.4	104.1	89.5	102.4

* 계절 요인들은 $ARIMA X-11$ 로 계산됨. CPS 실업률의 분모는 각각의 범주에 대해서 CPS 고용과 실직인원의 합이고, UI 지불요구율에서 분모는 CES 고용 총계임.

여기에서 실업의 계절적인 요인이 100보다 큰 값은 평균실업률보다 큰 달을 말한다. 신규 유입률은 겨울에는 평균실업률보다 낮고 여름에는 평균실업률보다 높다. 반면, 실직률은 반대의 형태를 보인다. 각 그룹에서의 큰 계절 실업률은 전체 실업률에 강한 영향을 미친다. 실직자와 신규 유입자는 경기순환과 다른 형태를 띤다는 점에서 수치적으로 중요하며, 전체 실업률의 약 반 정도를 설명한다. 이직자는 전체의 약 15%를 설명하므로 수치적으로 덜 중요하다. 늦여름과 가을에 구별되는 계절적 형태를 나타낸다. *UI* 지불요구율은 실직자의 계절적 형태를 따르나 신규 유입 또는 이직자의 계절적 형태를 따르지 않는다.

다음의 <표3.10>은 경기순환과 관련된 실업률 분포의 변화를 나타낸다. 경기후퇴 기간 중에는 노동력 요구는 떨어지고 실직자는 증가한다. 따라서 *UI* 지불요구율은 충분히 순환적인 지표가 된다. 그러나 *UI* 지불요구는 실직자의 순환적인 변화를 전적으로 반영하지는 못한다. 경기후퇴의 후반부 쪽에서는 실직기간이 길어져서 *UI* 보상혜택을 모두 써버린 노동자들이 증가한다. 또한 일단 재 고용되면 이러한 노동자들이 차후의 실직기간 동안 *UI* 보상 자격을 얻기 위한 고용신뢰도와 충분한 임금을 얻기까지는 얼마간의 시간이 걸린다. 중요한 사실은 실직자들에 대한 *UI* 범위를 살펴보면 오랜 기간에 걸쳐 지속적으로 줄어든 적이 있었다는 점이며, <표3.10>에서 이러한 사실을 확인할 수 있다.

Burtless et al.(1984)와 Corson et al.(1988)의 연구에 의하면 이러한 현상은 경제적인 변화나 인구 통계적인 변화와는 무관하며 *UI* 대상자들이 혜택을 적용 받지 못한 이유 때문으로 설명한다. 정책의 변화가 주요한 원인으로 주목된다.

이상의 토의는 대표적인 주의 움직임을 설명한 반면, 모형화 과정에서 설명되어야만 하는 중요한 주 간의 차이점들이 존재한다. *UI* 자료에서 보면 주의 적격요구 기준, 수혜기간, *UI* 적용범위에 대한 행정관례에서의 변동사항들이다. 순환적이며 계절적인 움직임에 있어서 실제적인 차이들은 회귀계수와 모형의 모수에 영향을 끼친다.

<표3.10> 실직범주들의 상대적인 크기(40개 주 평균)

년도	CPS 실업(%)					이직
	U.S. 실업률	CPS 실직자 UI지불요구	지불요구	실직	신규유입	
76	7.7	93.1	36.9	43.1	41.9	14.8
77	7.1	86.5	34.5	41.3	43.6	14.9
78	6.1	87.6	31.4	37.7	46.1	16.2
79	5.8	87.5	32.8	39.4	44.4	16.1
80	7.1	76.7	35.8	47.6	38.3	14.0
81	7.6	68.1	31.9	48.8	38.4	12.7
82	9.7	61.8	34.3	56.3	34.5	9.1
83	9.6	50.7	27.7	55.1	35.5	9.2
84	7.5	52.2	25.7	50.1	38.9	10.9
85	7.2	56.6	26.9	48.3	40.1	11.5
86	7.0	60.3	27.9	47.7	38.9	13.3
87	6.2	57.0	25.3	45.6	40.2	14.0
88	5.5	62.3	26.4	43.9	40.4	15.5

UI 적용에 있어서 변화량은 <표3.11>에 주어졌다. <표3.11>은 주별 전체 CPS실업자에 대한 UI 지불요구율의 연간 평균값들의 전체평균(%), 주 내에서의 CV값, 연간 평균값들 중 최소값과 최대값을 나타냈다.

비율모형에서 회귀성분의 일반적인 형태는 다음과 같이 주어진다.

$$\begin{aligned} \text{실업률} = & \beta_0(t) + \beta_1(t)(\text{지불요구율}) + \beta_2(t)(\text{인구 대 취업자비}) \\ & + \beta_3(t)(\text{신규유입률}) \end{aligned} \quad (3.17)$$

여기에서

$$\text{지불요구율} = \frac{\text{continued claims w/o earnings}}{\text{CES employment}} \times 100,$$

$$\text{인구 대 취업자비} = \frac{\text{CPS employment}}{\text{CPS population (16+)}} \times 100,$$

$$\text{신규유입률} = \frac{\text{CPS 취업 신규유입자}}{\text{CPS 신규유입자} + \text{CPS employment}} \times 100.$$

<표3.11> 주별 전체 CPS실업자에 대한 UI 지불요구자(1976-1987)

주	UI지불 요구	CV	최소	최대	주	UI지불 요구	CV	최소	최대
AL	25.5	23.7	17.5	37.5	MT	32.2	21.6	23.1	44.2
AK	57.1	21.2	43.1	83.0	NE	30.0	18.7	20.7	43.9
AZ	24.3	13.8	19.8	29.5	NV	34.7	19.7	25.3	48.2
AR	28.1	21.6	20.2	38.5	NH	26.4	17.4	19.2	34.7
CO	23.6	10.9	20.2	28.1	NM	24.2	10.5	20.4	28.1
CT	37.5	17.0	29.3	48.9	ND	31.1	14.7	24.3	36.7
DE	29.9	17.1	22.8	38.0	OK	27.5	21.7	19.4	39.0
DC	26.6	10.7	21.9	30.3	OR	37.6	16.7	28.8	48.4
GA	25.8	15.0	21.2	33.4	RI	51.2	12.7	39.5	61.4
HI	32.9	9.5	29.0	39.8	SC	26.3	18.1	19.8	33.9
ID	31.6	15.4	24.7	39.5	SD	22.5	30.6	14.7	35.0
IN	24.4	17.5	19.4	33.5	TN	27.2	26.2	18.1	41.0
LA	30.3	20.9	22.6	42.1	UT	31.4	25.7	19.9	41.1
KS	35.5	13.6	29.5	45.9	VT	40.0	9.7	33.9	46.3
KY	29.2	29.7	17.2	40.0	VA	17.6	20.1	13.3	24.7
LA	27.9	20.6	20.0	37.6	WA	36.1	15.7	30.1	49.1
ME	35.8	11.2	28.7	41.1	WV	32.8	24.7	21.1	42.7
MD	28.5	12.2	23.6	34.0	WI	34.2	24.1	24.8	47.1
MN	33.7	18.7	24.0	44.4	WY	28.7	28.7	20.7	49.7
MS	26.1	19.7	20.1	35.3	전체	31.0	23.2	17.6	57.1
MO	32.8	19.5	24.3	44.2					

지불요구율은 UI 혜택을 받고 있는 실직자들의 상대적인 규모를 나타내는 척도이다. 인구 대 취업자비는 실직보험 지불 요구인원에 포함되지 않은 실직자들을 나타내는 척도로 이용된다. 이직자의 수에 영향을 미치는 노동요구에 있어서의 계절적 변동을 인구 대 취업자비에서 찾을 수 있다. 즉 여름에 정점을 취하고 가을에 떨어지는 계절적 변동을 나타낸다. 이러한 주기 동안 수많은 계절적 직업들이 발생하고, 이직에 기인한 실직이 증가한다.

대부분의 주에서는 인구 대 취업자비에서 분모의 값을 CPS 취업자료를 이용하여 계산한다. 주의 CPS 취업자료는 표본오차의 영향을 받을 가능성이 있지만, CV값은 주의 CPS 추정값보다는 5~6배 정도 작다. 소수의 주에서 인구 대 취업자비의 계산에서 CES 자료를 이용한다. 대부분의

주들에서는 CES 자료는 CPS 취업축도와는 어느 정도 다른 계절적 형태를 갖고 있으며 실업률의 참값과는 상관성이 높지 않은 것으로 나타난다. 표본오차의 효과를 줄이기 위하여 CPS 신규 유입률 변수는 주보다는 훨씬 큰 지역(전국 또는 센서스 지역)에 대해서 계산된다. 몇몇 경우에 있어서는 주의 CPS 신규 유입률의 3-month 이동평균이 사용된다.

40개 주의 각각에 대한 모형들이 1976-1987년 동안의 월별 CPS 실업률의 시계열 자료에 적합되었다. 이 전에 언급한 것과 같이 CPS 자료는 다음과 같이 시그널 성분과 노이즈 성분으로 표현된다.

$$Y(t) = \theta(t) + e(t),$$

여기에서

$$\theta(t) = X(t)\beta(t) + u(t)$$

로 주어지고 계수들은

$$\beta(t) = \beta(t-1) + v_\beta(t)$$

와 같은 랜덤워크 과정을 따른다. $u(t)$ 와 $e(t)$ 의 효과는 분리하여 추정하지 않기 때문에 관측된 값들은

$$Y(t) = X(t)\beta(t) + w(t),$$

$$w(t) = u(t) + e(t) \sim ARMA(p, q)$$

와 같이 표현할 수 있다.

$S_w(t)$ 를 $\max(p, q+1)$ 과 같은 계(order)를 갖는 $w(t)$ 의 상태벡터라 하면, 상태공간 모형은 다음과 같은 전이방정식(transition equation)을 갖는다.

$$S(t) = \begin{pmatrix} \beta(t) \\ S_w(t) \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & T_w \end{pmatrix} \begin{pmatrix} \beta(t-1) \\ S_w(t-1) \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & \Gamma_w \end{pmatrix} \begin{pmatrix} v_\beta(t) \\ v_w(t) \end{pmatrix},$$

$$Y(t) = (X(t) \ 1 \ 0 \ \dots \ 0)S(t).$$

이 시스템의 모수들은 $Cov(v_\beta(t)) = Diag(\sigma_{\beta_1}, \dots, \sigma_{\beta_k})$, $Var(v_w(t))$ 이고, T_w 는 p AR모수들을 갖고, Γ_w 는 q AR모수들을 갖는다. 이러한 모수들은 우도함수의 새로운 형태를 이용하여 추정된다(Schweppe(1965)).

만약 백색잡음오차 $v_\beta(t)$ 와 $v_w(t)$ 가 정규분포를 따른다면, 1-step 앞의 예측오차들인 $\tilde{Y}(t)$ 는 독립인 $N(0, f(t|t-1))$ 을 따르는 확률변수가 된다.

관측된 표본오차의 결합확률은 각각의 밀도함수들의 곱으로 표현된다. 만약 상태벡터가 l 개의 비 정상원소를 갖고 있다면, 결합밀도함수는 처음 l 개의 관측점에 대한 조건부 함수로 표현되며, 미지인 모수들인 Ω 의 함수로써 로그우도함수

$$L(\Omega) = -\frac{1}{2} \left(\sum_{t=1}^n \log f(t|t-1) + \frac{Y(t)^2}{f(t)} \right)$$

는 상수의 범위 내에 있게된다. 만약 Ω 가 주어진다면, 초기값 $S(l+1|l)$, $P(l+1|l)$ 을 계산하기 위한 처음의 l 개의 관측값을 이용하여 $L(\Omega)$ 를 결정하는데, 이때 KF 반복법이 사용된다. 일반적으로 이러한 방법은 어려운 비선형 최적화 문제에 해당된다. 초기에 우리는 $Cov(v_\beta(t)) = qD$ 로 나타냈다. 여기에서 q 는 상수이고, D 는 사전에 명시된 상수들의 대각행렬이다. 또한 $w(t)$ 에 대한 1계인 AR 모형으로 출발하여, 두 모수들의 추정 문제를 다루었다. 계수들과 AR 모수 값의 변동의 정도에 대한 개략적인 추정값들을 계산하였고, 몇몇 경우에서 이러한 추정값들은 Watson과 Engle(1983)에 의해 개발된 EM -scoring 알고리즘의 초기값들로 이용하여 좀 더 다듬어지게 되었다. 이러한 알고리즘은 계수의 변화와 관측오차를 동시에 고려하는 일반적인 자기회귀 구조에도 이용된다.

모수들에 대한 초기값들은 합리적인 반복횟수 내에서 수렴값을 얻기 위해 매우 중요하다. 계수들의 표준편차 $\sqrt{\sigma_{\beta, \beta}}$ 는 관측오차에 대한 표준편차의 약 0.6%정도였다. 초기값들로 이러한 사실들을 이용한다면 EM 알고리즘은 일반적으로 3~6회의 반복 후 수렴하게 된다. Harvey와 Phillips(1979), Ansley와 Kohn(1985), Bell과 Hillmer(1987a)에 의하면 KF 를 초기화하는 많은 방법들을 소개하고 있다.

3.2 캐나다

3.2.1 서론

(1) 배경 및 목적

캐나다 노동력조사(LFS:Labour Force Survey)는 대규모 노동시장의 변화 양상 및 시의성 있는 노동시장의 정보를 파악하기 위해 2차 세계대전 이후 도입되었고, 주로 주(Province) 지역 및 국가 단위의 고용 및 실업통계를 생산할 목적으로 설계되었다. LFS는 1945년 분기별 조사로 시작하여 1952년 월별 조사로 변경되었고, 1960년부터 캐나다 실업통계를 생산하기 위한 공식조사로 승인되었다. 그 후 LFS를 통해 노동시장의 다양한 통계를 작성할 수 있도록 표본개편 및 조사방법에 관한 연구가 지속적으로 진행되었고, 현재는 캐나다 노동시장의 세부적인 변화에 관한 정보를 제공할 수 있을 정도로 발전을 거듭하였다. 매월 고용인구와 실업인구 총계 및 실업률에 관한 추정치, 노동인구의 특성(연령, 결혼여부, 교육정도, 가족현황) 등에 관한 공식통계는 LFS를 통해 작성된다.

LFS는 주 지역 및 전국 단위의 고용 및 실업통계 작성 외에 고용보험 경제구역(EIER:Employment Insurance Economic Regions), 센서스 도시 지역(CMA:Census Metropolitan Areas) 등과 같은 주 내의 특정 행정구에 대한 통계작성도 가능하도록 표본이 설계 되어있다. 최근에 들어서는 주 지역 내의 센서스 조사구(CD:Census Divisions)와 같은 소지역들에 대해서도 소지역 추정기법을 이용하여 관련 통계지표를 작성하고 있으며, 지방정부의 소지역에 대한 예산 배분 또는 정책 결정 등의 사안에 이러한 소지역 통계들이 이용되고 있다.

LFS 추정치들은 매월 "Labour Force Information"라는 책자를 통해 공표된다. 또한 노동시장의 좀 더 다양한 정보들은 캐나다 통계국의 전자 정보 데이터베이스의 일종인 "CANSIM"을 통해 획득할 수 있으며, LFS의 결과로부터 매월 9000 항목 이상의 시계열 자료들이 정기적으로 수정 보완된다. 이외에 노동시장의 중심지표가 되는 다양한 주제에 대한 세부적인 고찰을 다룬 "Labour Force Update"가 1997년부터 계간지로서 출간

되고 있으며, 1976년 이래로 최근까지의 방대한 시계열 자료(time series data) 및 횡단면 자료(cross-sectional data)를 포함하고 있는 “Labour Force Historical Review on CD-ROM”이 매년 제작되고 있다.

캐나다 노동력 조사에 의해서 매월 발표되는 통계수치는 자영업, 부업과 전업을 포함한 취업자 총수와 실업자 총수이다. 매월 발표하는 노동시장의 표준지표는 실업률, 취업률과 경제활동 참가율이고 노동력 조사의 주요 정보 요소로서 15세 이상 인구의 개인적 특성은 나이, 성별, 혼인상태, 교육정도와 가족사항이다.

취업통계의 추정값들에는 인구학적 특성, 산업과 업종, 정규직과 통상적인 근로시간 등이 포함되어 있으며 설문내용에는 비자발적 부업적 취업, 복수 직업 여부와 휴직 등에 대해서 분석할 수 있는 주제들이 포함되어 있다. 특히 1997년 이후에는 근로자들의 노조가입 여부와 임금수준에 대한 정보와 작업장의 근로자 수 및 직업의 정규직 또는 임시직 여부에 대한 정보를 제공하고 있다.

실업통계의 추정값은 인구학적 범주별, 실업기간, 구직활동 전의 활동 및 바로 이전 직장에서 이직한 이유 등에 대한 정보를 제공하고 있다. 노동력 조사에 의해서 발표되는 통계는 국가 단위와 주 단위 추정값이 핵심적인 내용이지만 경제구역(ER : Economic Region)과 센서스 도시지역(CMA ; Census Metropolitan Area)과 같은 소지역 단위에 대한 노동력 상태의 추정값을 제공하고 있다.

(2) 노동력 상태 결정과정

취업과 실업의 개념은 생산의 요소로서 노동력 공급이론을 근거로 정의하였으며 생산은 국민계정 체계(SNA : the System of National Accounts)에서 언급한 것과 같이 상품과 서비스로 정의되는 개념이므로 작업의 목적이나 성질에서는 보수를 받는 근로활동과 조금도 차이가 없는 무보수의 가사노동이나 자원봉사활동은 근로시간으로 계산하지 않고 있다.

노동력 공급의 측정단위는 개별적인 근로시간이지만 조사에서 모집단의 개별적인 구성원들의 구분은 취업, 실업, 비경제활동 인구로 분류되어야 한다. 조사기간 중에 보수 근로 중인 사람은 근로시간에 상관없이 취업자로 구분하고 근로시간과 무관하게 노동시장에서 구직 행위가 있을 경우에는 실업자로 구분한다. 나머지 인구는 현재 일을 하고 있지 않거나 또는 노동시장에서 구직 활동하지 않는 경우로 비경제활동 인구로 정의한다.

통계조사에서 적용하는 취업자와 실업자의 정의와 개념은 국제노동기구(ILO : International Labor Organization)의 기준에 준거하고 있다.

(가) 취업자(Employment)

조사대상 기간 중에 직업이 있거나 개인 사업에서 일을 하는 사람을 말하며 자기에게 직접적인 소득이 없을 지라도 가구단위로 운영되는 농장, 사업체 또는 전문적인 기관에서 일하는 경우도 포함된다. 또한 직업이나 사업체가 있을지라도 일시적인 병이나, 휴가, 노동쟁의 등의 이유로 일을 하지 못하는 일시적인 휴직자도 포함한다.

(나) 실업자(Unemployment)

이용되지 못하는 공급된 노동력으로 실업자를 정의하고 있으며 실업자의 구분은 구직행위와 근로행위의 준비여부를 기준으로 하고 있으나 구직행위는 가구조사에서 구체적이고 지속적인 의사 표명을 전제로 하고 있다. 실업자는 조사 대상 기간 중 다음 3가지 항목 중 하나에 해당하는 사람으로 정의될 수 있다.

- ① 현재 휴직 중이지만 근로 활동이 가능하고 복직을 기대할 수 있는 경우
- ② 일을 하지 않고 있으나 구직 활동을 하고 있으며 일이 주어지면 바로 시작할 수 있는 경우
- ③ 조사대상 기간으로부터 4주 이내에 새로운 일을 할 직업이 주어졌으며 바로 일을 할 수 있는 경우

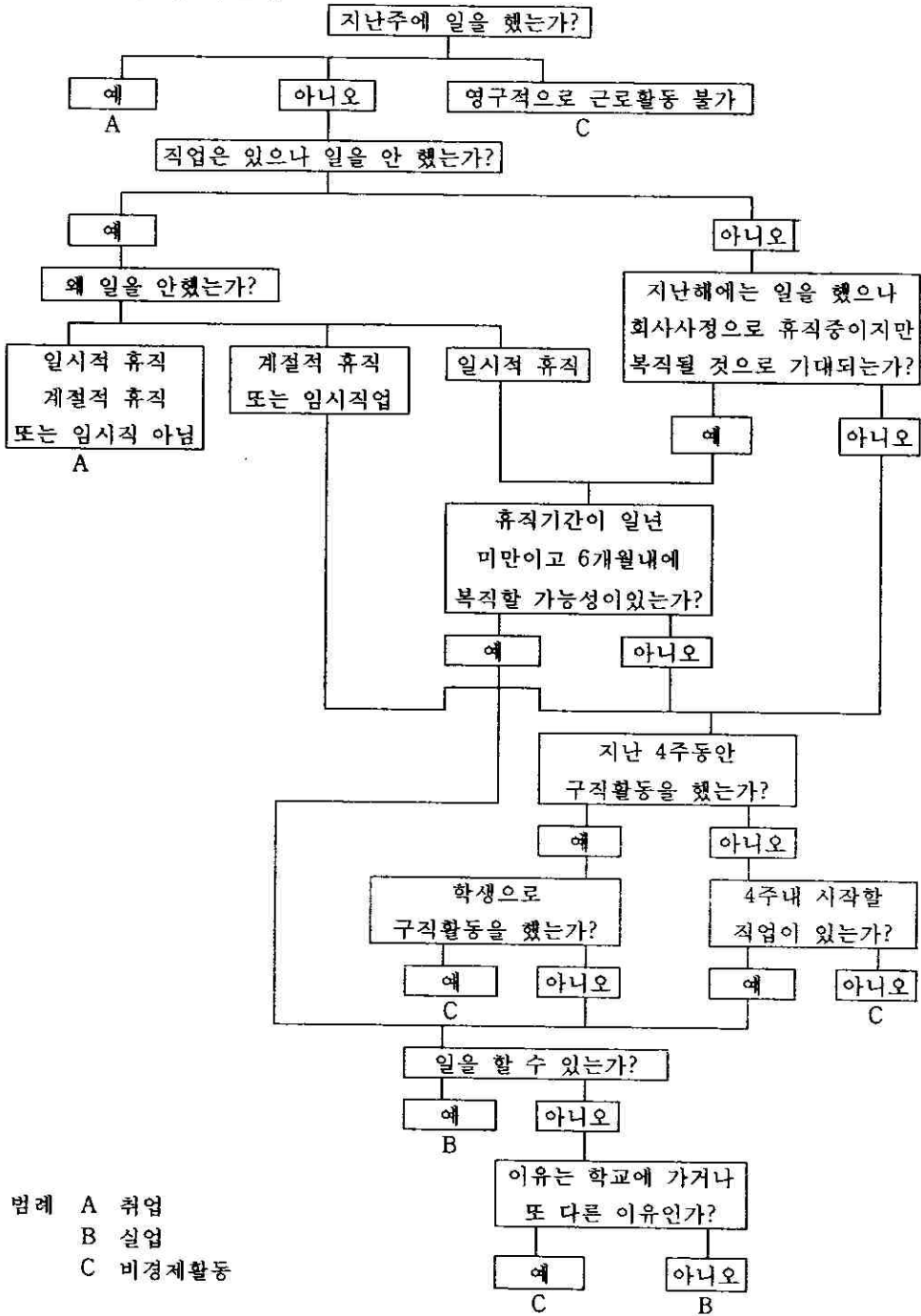
현재 학교에 다니는 학생이 전업적인 일을 찾고 있을 지라도 구직활동

으로 간주하지 않으며 여름방학동안의 근로 활동이나 구직 활동 역시 실업자의 분류로 고려하지 않고 있다.

(다) 비경제활동인구

노동시장에서 노동력을 제공하거나 참여할 의사가 없거나 가능성이 없는 사람을 말하며 이들은 실업자도 아니고 취업자도 아닌 상태의 사람들이다. 여기에는 순수하게 가사만 돌보는 사람, 학생, 투자배당금이나 연금 등을 받아서 생활하는 사람들이 포함된다.

■ 노동력 상태결정



LFS는 캐나다 인구의 약 98%에 해당하는 인구를 목표모집단(target population)으로 설정하였다. 캐나다 북서부지역, 인디언 보호구역, 왕실 소유지, 수감자 및 직업군인에 해당하는 약 2%는 LFS의 조사대상에서 제외된다. 1998년 12월 현재 LFS의 표본크기는 52,350가구이다. 1970년 표본개편을 통해 LFS의 표본크기는 이전의 35,000가구에서 50,000가구로 증편된 후, 1980년 47,000가구로 감소되었다가 1989년 주 지역 내의 행정 구역인 EIER 지역의 통계작성에 신뢰성을 확보하기 위해 63,000조사가구로 표본설계가 대폭 개편되었으며, 1993년에 약 59,000 조사가구로 다시 감소된 후 1995년부터 52,350 조사가구(EIER 지역에 16,500 조사가구, 나머지 지역에 35,850 조사가구)로 조정되어 현재에 이르고 있다.

LFS의 표본교체(sample rotation) 체계는 일종의 연동교체표본설계(rotating panel sample design)를 따른다. 표본가구는 6개의 부차표본(sub-sample)로 분리되어 6개월간 관리되며, 매월 1/6의 표본이 새로운 표본으로 대체되는 형식이다. 캐나다의 LFS는 매월 15일이 포함되는 주중에 실시되며, 80명의 조사전문인력을 포함하여 약 850명의 조사인력이 투입되어 조사가 이루어지고, 조사결과는 5개의 지방사무소(RO: Regional Office)에서 각각 취합되어 중앙으로 이관된다. 첫 달의 조사는 방문면접 형식을 취하며, 이 후 연속되는 다섯달은 전화조사를 통해 조사가 이루어진다. 조사자는 휴대용 컴퓨터를 이용하여 직접 설문 항목의 결과를 입력하는 일종의 컴퓨터 보조 면접(CAI: Computer Assisted Interviewing) 방식을 이용한다. 조사결과는 조사완료 시점에서 정확히 13일 후에 공표된다.

3.2.2 층화 및 추출단위 구성

캐나다의 주(Province) 지역들은 지리적인 경계에 의해 여러 개의 경제구역(ER:Economic Regions)들로 분할되며, 현재 캐나다 ER 지역은 총 72개가 존재한다. LFS에서는 1960년대 이후로 이러한 ER 지역들을 캐나다 노동력 조사의 1차 층(primary strata)으로 이용해오고 있다. 주 지역

및 전국 단위의 통계 작성에 ER 지역들이 이용된다.

초기의 LFS에서는 ER 지역이 표본설계 시 반영되었던 주 지역 내의 유일한 행정구역이었고 노동력 조사의 관심은 이러한 ER 지역에 집중되었으나, 1989년부터 HRDC(Human Resources Development Canada)의 자금지원에 의해 16,500개의 표본조사가가 LFS에 추가되면서 EIER 지역에 대한 노동력 조사도 추가적으로 가능하게 되었다. 따라서 현재의 노동력조사에서는 ER 지역과 EIER 지역 모두가 표본설계 시 층화에 반영되며, 추가표본은 주로 EIER 지역의 추정치의 신뢰도를 확보하기 위해 할당된다. 여기에서 ER 지역과 EIER 지역은 서로 조사 목적이 상이한 지역들이며 133개 지역이 서로 중복되어 조사된다. 한편 LFS 층화 시 반영되는 행정구역으로 CMA 지역을 들 수 있다. CMA 지역은 인구 100,000명 이상인 지역들로써 현재의 CMA 지역은 EIER 지역과 정확히 일치한다.

대영역 내에서의 세부적인 층화는 지리적인 구분에 관계없이 집락화 알고리즘에 의해 이루어진다. 그룹들 간의 가중 제곱합을 최소화하는 층화변수들을 이용하여 가능한한 동질적인 층으로 분할되며, 세부적인 알고리즘은 Drew et al.(1985)과 Singh et al.(1990)에서 참조할 수 있다. 층화 알고리즘에 이용되는 층화변수들은 다음과 같다. 이 층화변수들은 1991년 센서스 자료를 이용하여 선정되었으며, 각각의 층화변수들은 전체 인구의 2%이상을 설명할 수 변수들로 선정되었다.

- 농업부문 종사자 수
- 임업, 어업부문 종사자 수
- 광업부문 종사자 수
- 제조업(소비재분야) 종사자 수
- 제조업(고무, 플라스틱, 가죽분야) 종사자 수
- 제조업(섬유, 의류 분야) 종사자 수
- 제조업(가구, 펄프, 제지, 인쇄, 목재분야) 종사자 수
- 제조업(금속, 광업분야) 종사자 수

- 제조업(석유화학, 화학분야) 종사자 수
- 운수업 부문 종사자 수
- 건설업 부문 종사자 수
- 서비스업(상업분야) 종사자 수
- 서비스업(금융분야) 종사자 수
- 서비스업(개인/사업분야) 종사자 수
- 서비스업(정부분야) 종사자 수
- 종사인원 총계
- 총 소득
- 15세 이상 인구
- 15-24세 인구
- 55세 이상 인구
- 1인 거주 가구 수
- 2인 거주 가구 수
- 개인 소유 가구 수
- 총 임차료
- 고졸학력 인구
- 영어를 모국어로 하는 인구
- 프랑스어를 모국어로 하는 인구
- 영어/프랑스어 이외의 언어를 모국어로 하는 인구

LFS 추출틀은 농촌 지역, 인구 50,000명 이상의 대도시 지역과 소도시 지역의 3가지 유형의 지역들로 구분된다. 각 지역에 대한 총화는 다음과 같은 방법으로 이루어진다.

농촌지역에서 총화는 EI 지역과 EIER 지역의 교집합 내에 있는 2~3개의 CD(Census Division) 지역들을 함께 묶어 지리적 층으로 구성하였다.

캐나다 내의 대도시 지역인 17개의 CMA 지역들에 대해서는 충분한 수의 아파트들이 있기 때문에 각각의 CMA 지역들에 대해 독립적인 아파

트 추출틀을 작성하며, 아파트 추출틀을 제외한 나머지 지역에 대해서는 일종의 지역 추출틀(area frame)을 형성하였다. 또한 \$100,000 이상의 평균소득을 갖는 고소득 지역들은 독립된 층으로 구성하여 고소득 가구들에 대한 대표성을 제고하였고, 부수적으로 소득관련 조사 및 소득관련 정보 수집이 용이하도록 하였다. 이러한 부수적인 정보는 고소득층에 대한 무응답의 경향을 분석하고자 하는 경우에도 유용하게 이용될 수 있다. 아파트 추출틀과 고소득 층을 제외한 나머지 지역들은 SNF(Street Network File) 지역으로 구분하였다. 최종적으로 마지막 층에는 적어도 48가구의 표본이 배정되도록 하였다.

LFS 표본설계에서 대규모 CMA 지역들에 대해서는 아파트 추출틀을 사용해오고 있다. 현재 18개 CMA 지역에서 표본 추출틀로써 이 목록이 이용되며, 각 CMA 지역에서 새로운 아파트가 건설되면 바로 표본목록에 추가된다. 또한 7개의 CMA 지역에 대해 평균소득이 \$20,000 미만의 저소득 아파트 단지를 파악하여 저소득 아파트 층을 구성하여, 고소득 아파트 층과 더불어 소득관련 정보를 획득한다. 캐나다의 각 CMA 지역에 대한 아파트 추출틀 층화 현황은 다음 <표3.12>에 주어졌다.

<표 3.12> 아파트 추출틀 층화

CMA	지리적 층	층의 총수	CMA	지리적 층	층의 총수
Halifax	2	2	London	1	2
Quebec	2	2	Windsor	1	2
Montreal*	4	9	Winnipeg*	1	6
Ottawa-Hull*	3	6	Saskatoon	1	1
Oshawa	1	2	Calgary*	1	3
Toronto*	6	16	Edmonton*	1	3
Hamilton	2	4	Vancouver*	4	6
St. Catharines	1	1	Victoria	1	1
Kitchener	2	2	Total	34	68

- (주) ① “*” 는 저소득 아파트 층을 갖는 CMA 지역을 표시함
 ② “층의 총수”는 저소득 아파트 층을 포함한 수임

대부분의 소도시에서는 EA(enumeration area) 지역을 층화단위로 이용하였다. 조사 비용을 절약하기 위해 각 층에서 직접 표본가구를 추출하지 않고, 우선 각 층을 여러 개의 집락으로 구분하고 추출된 집락 내에서 표본가구들을 추출하였다. 통상적으로 농촌지역에서는 EA 지역이 집락으로 사용되었고, 도시지역에서는 좀 더 다양한 형태의 집락들이 이용되었다. 다음 <표3.13>은 LFS 표본설계에 이용된 일 단계 단위(first-stage unit)의 유형들을 요약한 것이다. 여기에서 “추출가구 수”는 LFS에서 조사되는 가구 수를 나타낸다. 집락과 표본가구에 대한 추출방법은 다음 절에서 논의하기로 한다.

<표 3.13>일단계 단위(first-stage unit), 단위당 가구 수와 추출가구 수

지 역	추출단위	단위당 가구수	추출가구수
Toronto, Montreal, Vancouver	cluster	200-250	6
Other cities	cluster	150-200	8
Apartment frame	apartment	varies	5
Most rural areas and non-SNF parts of cities	EA	300	10

3.2.3 표본배정, 추출, 순환

(Sample allocation, selection and rotation)

LFS의 표본설계는 캐나다 전체, 주(Province) 지역, EIER 지역, CMA 지역, ER 지역에 대해 각각 다음과 같은 실업자 추정값의 목표 CV값이 만족되도록 설계되었다. 캐나다 전체의 실업자 추정치의 목표 CV 는 약 2%이내, 주(Province) 지역의 CV 값은 약 4~7%선에서 관리되도록 하였다. EIER 지역과 CMA 지역은 분기 실업자 추정치의 CV 값이 15% 이내가 되도록 하였고, 여기에서 하나의 EIER 지역에 배당되는 최소 표본크기는 매월 600가구가 배정된다. ER 지역에 대한 분기별 목표 CV 값은 25% 이내가 되기를 기대하고 있다. 캐나다의 각 주 내의 ER 지역, EIER

지역과 CMA 지역의 현황은 다음 <표3.14>와 같다.

LFS의 표본크기는 총 52,350가구로써 HRDC의 자금지원에 의해 추가된 16,500 표본가구를 제외한 35,850 표본가구는 전국 및 주(Province) 지역의 최적 추정을 위해 배정되며, 이 표본을 핵심표본(core sample)이라고 지칭한다. 핵심표본은 각 주의 실업률 추정의 목표정도를 만족하도록 배정된다. 주 내에서 각각의 ER 지역에 대한 핵심표본의 배정은 총 가구수에 비례하여 배정하며, 최소한 200~300 표본가구가 배정되도록 하였다. 또한 16,500 가구의 추가표본은 핵심표본을 보충하여 EIER 지역의 추정의 정확도를 높이기 위해 추가로 배정된다. 우선 핵심표본 만으로 EIER 지역에 대한 CV 값을 계산 한 후, CV 값이 큰 EIER 지역에 대해 추가 표본을 배정하는 형식을 취하며, EIER 지역에 대해서는 최소한 600 가구가 배정되도록 하였다.

<표 3.14> ER, EIER, CMA 지역 현황

주(Province)	ERs	EIERs	CMAs
Newfoundland	4	3	1
Prince Edward I	1	1	0
Nova Scotia	5	5	1
New Brunswick	5	4	1
Quebec	16	13	6*
Ontario	11	18	10*
Manitoba	8	3	1
Saskatchewan	6	4	2
Alberta	8	4	2
British Columbia	8	6	2
Canada	72	61	25*

(주) Ottawa-Hull CMA 는 Ontario와 Quebec 양쪽에서 카운트 됨.

LFS에서 표본 추출은 대부분의 지역에서 이 단계 추출(two-stage sampling) 만을 이용한다. 일 단계 추출은 지역 추출(area sample)이고,

이 단계 추출은 일 단계 추출 지역들에 대해 거주가구 목록을 작성하여 이 목록으로부터 표본가구를 추출한다.

농촌지역의 경우 일 단계 추출단위는 EA지역, 이 단계 추출단위는 가구 단위가 된다. 일 단계의 지역 추출은 확률비례계통추출법, 이 단계의 가구 추출은 계통추출법이 이용된다. 아파트 추출틀을 이용하지 않는 주요 도시지역의 경우 일 단계 추출단위는 EA지역, 이 단계 추출단위는 가구 단위가 된다. 일 단계의 지역 추출은 Rao-Hartley-Cochran(RHC)의 랜덤그룹법이 이용되며, 이 단계의 가구 추출은 계통추출법이 이용된다. 아파트 추출틀을 이용하지 않는 기타 도시 지역들도 몇몇 특별한 경우를 제외하고는 주요 도시지역과 유사한 방법을 취하고 있다. 랜덤그룹법은 인구유입 및 유출양상을 LFS의 표본크기에 반영시킨 방법으로써 시기에 따라 유동성이 큰 대도시 지역의 인구변화를 표본크기에 반영시킨 방법이다. 아파트 추출틀을 이용하는 주요 도시지역의 경우 일 단계 추출단위는 아파트 단지, 이 단계 추출단위는 가구 단위가 된다. 일 단계 아파트 단지에 대한 추출은 확률비례계통추출법, 이 단계 가구단위의 추출은 계통추출법이 이용된다.

LFS에서는 표본의 일부가 매달 새로운 표본으로 교체된다. 다 단계 표본설계의 매 단계에서 표본 단위의 교체가 이루어지며, 표본추출의 최종 단위인 가구는 6개월마다 교체된다. 조사자의 업무 부담과 응답가구가 오랜 기간 동안 표본으로 조사되는 동안 발생할 수 있는 무응답에 대한 가능성을 최소화하기 위해 매 달 1/6의 표본이 대체된다. 따라서 하나의 집락에 포함된 표본가구는 연속적으로 6개월 간 조사된 후 표본에서 완전히 삭제되고 새로운 표본으로 대체된다.

3.2.4 특별조사와 보충조사

캐나다 노동력조사 외에 추가적인 많은 조사가 LFS 추출틀 또는 LFS 표본을 이용하여 실시되며, 이러한 조사들은 캐나다 정부 부처의 자금 지원에 의해 시행된다. 다음 <표3.15>은 LFS 연동교체표본 또는 LFS

추출틀을 이용한 특별조사 및 보충조사 현황을 요약한 것이다.

<표3.15> 특별조사 및 보충조사 현황(1998년 현재)

Survey	조사기간	Survey	조사기간
Canadian Travel Survey	1-12월 (매월조사)	Homeowner Repair and Renovation Survey	3월
Employment Insurance Coverage	1월	Survey of Consumer Finances	4월
Survey of Household Spending	1월-3월	Cultural Capital Survey	4월
Survey of Labour and Income Dynamics	1월, 5월	National Population Health Survey	2, 6, 8, 11월
Adult Education and Training Survey	1월	National Longitudinal Survey of Children	11월
Resident Telephone Services Survey	2, 5, 8, 11월	Survey of Work Arrangements	11월
Survey of Household Energy Use	2월		

SCF(Survey of Consumer Finances)는 일년에 한번 시행되는 소비자 재정에 관한 조사로써 보통 4월에 실시된다. 모든 가구들이 4개의 순환 그룹에 배정되어 LFS 조사에 추가된다. 4월의 LFS 조사에 앞서 각 가정에 우편으로 설문지가 배달되고 LFS 조사 기간 동안 회수된다. SCF에서 조사하는 주요 정보는 소비자의 평균 소득과 세금 공제 전과 공제 후의 소득관련 정보들으로써, 이러한 결과들은 저소득 기준점 결정 등과 같은 소득 관련 측도들로 이용된다.

SHS(Survey of Household Spending)는 일년에 한번 실시되는 가계비 지출 및 식료품 지출 조사로써 보통 1월~3월에 걸쳐 시행되며, 소비자 가격 지표 산정의 정보로써 이용된다. SHS 조사는 LFS 표본을 포함하는 집락들에서 표본가구를 추출하나 표본가구들은 LFS 조사와는 별도로 조사된다.

SLID(Survey of Labour and Income Dynamics)는 노동력과 소득 변천과정 파악을 조사의 목적으로 일년에 2번, 1월과 5월에 조사가 이루어

지며 LFS와 병행하여 시행된다. 조사 결과는 저소득 층의 유입, 유출 동향, 노동 시장의 변화, 가족변화와 경제적인 복지와의 상관관계 등을 분석하기 위해 이용된다.

NPHS(National Population Health Survey)는 일종의 국민 건강조사로써 분기별로 실시되며 국민 건강의 계절적 요인을 파악하기 위해 NPHS 표본을 2월, 6월, 8월과 11월조사에 각각 1/4씩 배분한다. NPHS 조사는 주 정부의 자금지원에 의해 실시되며, LFS 조사와 병행하여 실시되지는 않는다. 초기에 선택된 가구의 구성원은 2년에 한번 심층 면접을 받으며 20년 동안 지속적으로 관리된다. 기초적인 건강정보는 거주 가구의 모든 구성원들에 대해서 취합되며 이러한 시계열 자료는 횡단면 추정 목적에 이용된다.

NLSCY(National Longitudinal Survey of Children and Youth)는 아동에 대해 유아기부터 성인기까지의 발달 과정을 모니터하는 조사로써 일년에 한번 실시되는 일종의 시계열적 장기조사에 해당한다. LFS 조사가구의 약 30%만이 대상연령에 있는 아동을 포함하는 관계로 NPHS의 추가 표본이 조사에 이용되며, 조사방법은 NPHS 조사와 유사하다.

3.2.5 가중치와 추정

LFS의 가중치는 다음의 세가지 요인들에 기인하여 작성된다. 표본설계를 반영하는 일종의 설계 가중치, 무응답 가구를 보정하는 일종의 무응답 보정 가중치와 모집단 총계에 표본 추정치를 일치시키는 일종의 사후 총화 가중치(g-factor)의 3가지 요인들에 의해 LFS의 최종 가중치가 결정된다. LFS 조사 추정치는 LFS 표본이 확률표본이기 때문에 추정치의 표본오차를 추정하여 신뢰도를 판단할 수 있다. 표본설계에 계획되지 않은 관심지역들에 대한 추정문제는 소지역 추정기법을 도입하여 추정량의 신뢰도를 확보하였다. EIER 지역들이 여기에 해당된다.

LFS에서는 가구단위에 대해 계통추출이 이루어지는 마지막 단계를 제외하고는 모든 단계에서 확률비례추출법으로 표본이 추출되는 총화 다단

계 추출법이 이용된다. 설계가중치는 이러한 복합표본설계에서 가장 기본적인 가중치로써 각 추출단위에 대해서 추출률의 역수로 결정된다. 예를 들어 층에 대한 설계가중치는 $R_h = N_h/n_h$ 와 같이 표현될 수 있다. 여기에서 n_h 는 층 h 에 대한 표본가구 수, N_h 는 표본설계 시 층 h 에 있는 총 가구 수를 나타낸다.

이 단계 추출의 경우는 다음과 같이 설명될 수 있다. n_{hj}^* 를 층 h 에 있는 j 번째의 일 단계 추출단위(FSU: First Stage Unit)라고 하자. 추출해야 할 FSU의 수는 $n_{1h} = n_h/n_{hj}^*$ 로 주어진다. h 층에서 j 번째 FSU에 있는 가구들의 수를 N_{hj} 라 하면 j 번째의 FSU에 대한 추출율은 $R_{hj} = N_{hj}/n_{hj}^*$ 로 주어지며, 이때 j 번째 FSU에 대한 일 단계 산입확률(inclusion probability)은

$$\pi_{1hj} = \frac{n_{1h}}{\sum_{j \in h} R_{hj}} R_{hj}$$

로 나타낼 수 있다. j 번째 FSU가 주어진 상태에서 k 번째 가구가 선택될 조건부 산입확률은

$$\pi_{k/j} = \frac{n_{kj}^*}{N_{hj}} = \frac{1}{R_{hj}}$$

로 주어지며, h 번째 층에서 k 번째 가구에 대한 산입확률은

$$\begin{aligned} \pi_{hk} &= \pi_{1hj} \pi_{k/j} \\ &= \frac{n_{1h}}{\sum_{j \in h} R_{hj}} R_{hj} \frac{1}{R_{hj}} \\ &= \frac{n_{1h}}{\sum_{j \in h} R_{hj}} \end{aligned}$$

로 계산될 수 있다. 여기에서

$$\begin{aligned} \sum_{j \in h} R_{hj} &= \sum_{j \in h} \frac{N_{hj}}{n_{hj}} \\ &= \frac{n_{1h}}{n_h} \sum_{j \in h} N_{hj} \\ &= n_{1h} R_h \end{aligned}$$

인 관계가 성립하며, 결국 산입확률은 $1/R_h$ 과 같게 된다. 참고적으로 LFS에서는 기본적으로 각 층 내에서 동일한 설계 가중치 R_h 를 이용한다.

LFS와 같은 연속조사에서는 고정된 추출률을 유지할 경우, 인구 증가 및 인구 유입 등으로 시간이 지남에 따라 표본 수도 증가하게 된다. 급속한 인구 증가가 발생하는 특정지역의 표본 수는 급격히 증가되어 조사원의 업무부담을 가중시키고 조사의 질을 떨어뜨릴 가능성이 있다. 이러한 문제점을 해결하기 위해 LFS 표본 설계에서는 이러한 집락들에 대해서 표본을 부차추출(subsampling)하여 전체 표본크기를 일정하게 유지한 후 집락 부차가중치(cluster subweight)를 산출하여 LFS 설계가중치에 추가하여 사용하였다. 구체적으로 SC(Subclustering)방법, SRC(Self-Representing Cluster)방법, CSS(Cluster Subsampling)방법 등을 이용한다.

SC 방법은 인구유입 등으로 인해 원래의 집락의 크기가 약 3배를 초과할 때 이러한 집락들을 여러 개의 작은 집락으로 분할하여 표본 가구를 추출하는 방법이다. 분할된 작은 집락들의 표본 수를 n_{2hj} 라 하자. N_{hji} 를 새로운 집락의 크기, n_{hji} 를 새로운 집락의 표본크기, R_{hji} 를 새로운 집락들에 대한 추출률이라 할 때, 새로운 부차집락들로부터 집락추출률

$$R_{hj}^* = \sum_{i \in j} \frac{R_{hji}}{n_{2hj}}$$

를 얻을 수 있다. 이때 집락 부차가중치는 $K = R_{hj}^*/R_{hj}$ 로 주어지며, 설계 가중치는 원래의 가중치에 이러한 부차가중치를 곱하여 계산된다.

SRC 방법은 임의의 층에서 증가하는 거주단위들의 특성이 나머지 단

위들의 특성과 구별되거나 집락의 규모가 층 규모의 20%를 초과할 경우, 해당 집락을 하나의 층으로 재분류하여 가중치를 결정하고, 나머지 집락들에 대해서는 가중치를 재 보정해 주는 방법이다. 이렇게 생성된 새로운 층을 (h_j) 라 하자. 이러한 층 내에서 새로운 집락들이 구성되어 표본들이 추출된다. $N_{(h_j)}^N$ 을 새로운 층의 크기, $n_{(h_j)}^N$ 을 새로운 층의 표본크기라 할 때, 층 추출률은 $R_{(h_j)}^N = N_{(h_j)}^N/n_{(h_j)}^N$ 으로 주어지며, 이러한 새로운 층으로부터 추출된 가구들은 가중치 $K = R_{(h_j)}^N/R_h$ 가 배정된다. 나머지 집락들의 가구들에 대한 가중치는 다음과 같은 방법으로 보정된다. 하나의 층에서 6개의 집락들이 추출되었다고 하자. 층 내에서 1개의 성장집락이 발생하여 새로운 층으로 재 분류 되었다면 나머지 5개의 집락의 가구들에 대한 가중치는 보정되어야만 한다. $N_h^R = N_h - N_{(h_j)}^N$ 를 새로운 층을 제외한 층의 크기, $n_h^R = n_h - n_{h_j}$ 를 표본의 크기라 할 때, 층에 대한 추출률은 $R_h^R = N_h^R/n_h^R$ 로 계산되며, 이때 집락 부차가중치는 $K = R_h^R/R_h$ 로 결정된다.

CSS 방법은 집락단위가 부차표본으로 추출되는 경우에 이용된다. 임의의 집락이 추출률 R_{h_j} 로 추출되고, 부차추출이 추출률 $R_{h_j}^*$ 로 추출되었다면, 이때 집락 부차가중치로써 $K = R_{h_j}^*/R_{h_j}$ 를 이용하는 방법이다.

표본추출의 마지막 단계는 계통추출이 이용된다. 가구에 대한 추출률은 일관되게 적용되기 때문에 인구 증가로 인한 표본가구수의 증가는 조사규모 및 조사비용의 증가를 초래한다. 이러한 조사비용을 통제하기 위해 표본 수 안정화 작업이 수행된다. 표본 수 안정화 작업은 전체 표본가구 수를 적절한 수준에서 유지하기 위해 초과 표본들을 랜덤하게 제거하는 작업을 말한다. 이러한 과정에서 가구단위의 산입확률(inclusion probability)은 당연히 바뀌게 된다. 현 표본설계에서는 같은 EIER 지역에 속해 있고 같은 순환그룹에 포함되어 있는 모든 가구단위들을 안정화 지역(stabilization area)으로 정의하고, 각 안정화 지역 a 에 대해서 표본크기

가 결정된다. 안정화 지역 a 의 표본크기를 b_a 라고 나타내자. 안정화 지역에서 안정화작업을 거치지 않은 표본크기를 n_a 라 할 때, n_a 가 b_a 를 초과한다면 $n_a - b_a$ 만큼의 표본가구가 랜덤하게 제거된다. LFS에서는 임의의 집락이 부차추출 되었을 경우에는 이 집락을 안정화 작업에서 제외시키기 때문에 이러한 집락에 포함된 가구단위들은 안정화 가중치(stabilization weight)의 영향을 받지 않는다. 이러한 집락을 제외시킨 안정화 지역 a 의 가구단위의 총계를 c_a 라 할 때, 지역 a 에 있는 조사가구의 안정화 가중치는 $s_a = (n_a - c_a) / (b_a - c_a)$ 가 이용된다.

통계조사에서 무응답은 항목 무응답(item nonresponse)과 단위 무응답(unit nonresponse)의 두 가지 유형으로 분류된다. LFS에서는 항목 무응답은 대체법(imputation)으로, 단위 무응답은 전체적인 가중치 조정을 통해 처리하고 있다. 항목 무응답은 지리적으로 또는 인구 통계적으로 유사한 특성을 갖는 응답자의 응답패턴을 이용하여 대체된다. LFS에서는 단위 무응답을 처리하기 위한 무응답 층을 “같은 EIER 지역에 속하고, 같은 유형의 지역적 특성을 가지며, 같은 표본순환그룹 내에 있는 가구들”로써 정의한다. 무응답 층을 올바르게 구성하였을 경우 동일한 무응답 층에 속한 응답 가구와 무응답 가구는 서로 비슷한 속성을 갖기 때문에 응답 가구가 무응답 가구를 대표한다고 가정할 수 있게 된다. 단위 무응답에 대한 보정은 설계 가중치에 보정요인

$$f_b = \frac{\sum_{k=1}^n \pi_k^{-1}}{\sum_{k=1}^r \pi_k^{-1}}$$

을 곱하여 계산한다. 여기에서 π_k^{-1} 은 각 표본가구에 부여된 설계 가중치, n 은 무응답 층 b 에 있는 표본가구 수, r 은 응답가구 수를 나타낸다.

LFS에서의 최종 가중치는 특별한 경우를 제외하고는 설계 가중치와 보조정보로부터 얻게 되는 사후 가중치(g-factor)의 곱으로 계산되며, 여

기에서 사후 가중치 계산은 일반적인 회귀추정방법이 이용된다. 자세한 추정절차는 Lemaitre and Dufour (1987)에 소개되어 있다. 먼저 다음과 같은 기호를 정의하자.

$p=1,2, \dots, 10$: 주 지역을 나타내는 기호,

$u=1,2, \dots, U$: p 번째 주 내에 있는 EIER 지역,

$f=1,2, \dots, F$: u 번째 EIER 지역 내에 있는 추출틀의 형태,

$h=1,2, \dots, H$: 추출틀 f 내에 있는 층,

$r=1,2, \dots, 6$: h 번째 층 내에 있는 순환그룹(패널),

$j=1,2, \dots, J$: r 순환그룹의 집락,

$k=1,2, \dots, K$: 집락 j 에서의 가구,

$i=1,2, \dots, c_k$: k 번째 가구 내에 있는 구성원,

SC(subclustering) 방법은 우선 해당 집락을 여러 개의 작은 집락들로 재 구성한 후 표본 추출을 위한 집락들을 선정하고 전체 표본 수를 참고 하여 선정된 집락 내에서 표본가구를 추출한다. 원래의 집락 추출률이 $R_{pufh \cdot j}$, 해당 집락 추출률이 $R_{pufh \cdot j}^*$ 라 하면 해당 집락의 부차 가중치는 $c_{pufh \cdot j} = R_{pufh \cdot j}^* / R_{pufh \cdot j}$ 이다.

SRC(self-representing cluster) 방법은 층 내에서 성장 집락을 분리하여 새로운 층(h)으로 구성하고 h 층 내에서 여러 개의 집락들을 재 구성한 후 표본가구를 추출하는 방법이다. 원래의 층 추출률이 R_{pufh} , 새로 형성된 층의 추출률이 R_{pufh}^* 라 하면 이때 새로 형성된 층에서 가구단위들에 배정되는 집락 부차가중치(cluster subweight)는 $c_{pufh} = R_{pufh}^* / R_{pufh}$ 이다. 성장집락을 제외한 나머지 집락의 가구단위에 대해서는 보정된 집락 부차가중치 $c_{pufh} = R_{pufh}^R / R_{pufh}$ 를 적용한다. 여기에서 R_{pufh}^R 은 나머지 층의 추출률을 나타낸다.

CSS(cluster Subsampling) 방법은 추출된 가구단위들이 부차추출되고 이러한 부차추출된 가구단위들만 조사하는 방법이다. 원래의 집락 추출률

이 $R_{pufh.j}$ 이고 부차추출하기 위한 해당 집락 추출률이 $R_{pufh.j}^*$ 라면 이때 집락 부차가중치는 $c_{pufh.j} = R_{pufh.j}^*/R_{pufh.j}$ 이다.

안정화 가중치(stabilization weight)는 앞서 언급되었던 안정화지역(stabilization area) 내에서만 적용된다. 각 안정화 지역 내에서의 표본크기를 $b_{pu...r}$, 실제 추출된 표본크기를 $n_{pu...r}$, 안정화 지역에서 CSS 방법으로 추출된 표본의 크기를 $c_{pu...r}$ 이라 할 때, 안정화 가중치는

$$s_{pu...r} = \frac{n_{pu...r} - c_{pu...r}}{b_{pu...r} - c_{pu...r}}$$

로 계산된다.

이상의 가중치들을 이용하여 조사가구에 대한 설계가중치를 산출하며 LFS에서는 다음과 같은 설계가중치를 고려한다.

$$\pi_{pufhrjk}^{-1} = w_{pufh} \times c_{pufh.j} \times s_{pu...r}$$

여기에서 w_{pufh} 는 같은 층 내에 있는 모든 가구단위들에 대해서 동일한 가중치를 배정했던 표본설계 당시의 가중치를 나타낸다. 표기상의 편의를 위해 앞으로 $\pi_{pufhrjk}^{-1}$ 를 π_k^{-1} 로 나타내기로 한다.

LFS에서는 무응답 층에 대해서 무응답 보정을 실시하며, 보정 가중치로써

$$f_{puf.r} = \frac{\sum_{k \in s} \pi_k^{-1}}{\sum_{k \in r} \pi_k^{-1}}$$

을 이용한다. 여기에서 분자의 s 에 대한 합은 무응답 층에 있는 모든 가구들에 대한 합을 나타내며, 분모의 r 에 대한 합은 무응답 층에 있는 모든 응답가구들에 대한 합을 나타낸다. 같은 무응답 층에 속해 있는 모든 가구단위들은 동일한 무응답 가중치를 갖는다.

무응답 보정요소가 추가될 경우 부차 가중치는 $a_k = f_{puf.r} \times \pi_k^{-1}$ 과 같이 설계가중치와 무응답 보정가중치의 곱으로 주어진다. 즉 같은 조사

가구 내의 모든 구성원들은 동일한 부차가중치를 갖는다.

위에서 언급한 부차가중치를 이용하여 고용 인구 Y 에 대한 총계 추정 값을 산출해 보자. 모집단에서 고용인구의 총계를

$$t_y = \sum_U y_i$$

라고 하자. 여기에서 U 에 대한 합은 모집단에서 관심영역 내에 있는 모든 구성원들의 합을 나타내며, y_i 는 조사대상자가 고용일 경우 1, 아닐 경우 0의 값을 갖는다. 이때 표본조사에 의한 총계 추정치는 부차가중치에 의존하며

$$\hat{t}_{ys} = \sum_s y_i a_i$$

와 같이 표현될 수 있다. 여기에서 s 에 대한 합은 표본으로 추출된 조사 대상자들에 대한 합을 나타내고, a_i 는 부차가중치를 의미한다. 위의 t_y 와 \hat{t}_{ys} 는 각각 다음과 같이 다시 표현할 수 있다.

$$t_y = \sum_{k=1}^N \sum_{i=1}^{c_k} y_i = \sum_{k=1}^N y_k,$$

$$\hat{t}_{ys} = \sum_{k=1}^n a_k \sum_{i=1}^{c_k} y_i = \sum_{k=1}^n y_k a_k.$$

여기에서 c_k 는 k 번째 조사가구의 구성원의 수, N 은 모집단의 가구 수, n 은 표본 가구 수, y_k 는 $y_k = \sum_{i \in k} y_i$ 를 의미하며, k 는 가구 총 수, i 는 구성원을 나타낸다.

LFS에서 적용하고 있는 마지막 단계의 가중치로써 사후층화 가중치 (g-factor)를 들 수 있다. 사후층화 가중치는 사후층화를 통한 보조정보로부터 획득하며 회귀추정방법을 이용하여 산출한다. 각 주 단위의 성별-연령대별 그룹, ER 지역과 CMA 지역에 대한 인구 총계, 센서스 결과에 의한 인구 추계 정보 등이 보조정보로 활용되었다. 추가적인 논의를 위해

다음과 같은 기호를 정의하기로 하자.

y_i : i 번째 조사자에 대한 특성치,

y_k : k 번째 가구에 대한 특성치 총계,

Q : 추정에 이용된 보조변수의 수, $q=1, 2, \dots, Q$,

x_{qi} : 조사자 i 에 대한 q 번째 지표변수의 값, 지표변수는 조사자 i 가 j 번째 범주에 속할 경우 1, 기타 0의 값을 갖는다.

x_{qk} : k 번째 가구단위에 속하는 조사자들에 대한 q 번째 지표변수값의 총계,

x_k : q 번째 원소가 x_{qk} 인 $Q \times 1$ 벡터,

c_k : k 번째 가구의 크기,

\hat{t}_{yq} : 위에서 언급한 부차가중치에 근거한 추정치,

\hat{t}_{xq} : q 번째 보조변수에 대한 부차가중치에 근거한 추정치.

사후층화 가중치를 산출하기 위해

$$\hat{t}_y = \hat{t}_{yq} + \sum_{q=1}^Q \hat{B}_q (t_{xq} - \hat{t}_{xq})$$

와 같은 회귀추정량을 이용한다. 여기에서

$$\begin{aligned} \hat{t}_{xq} &= \sum_s x_{qs} a_s, \\ \hat{B} &= (\hat{B}_1, \dots, \hat{B}_Q)^T \\ &= \left(\sum_{k=1}^n \frac{x_k x_k^T a_k}{c_k} \right)^{-1} \sum_{k=1}^n \frac{x_k y_k a_k}{c_k} \end{aligned}$$

이며, $\left(\sum_{k=1}^n \frac{x_k x_k^T a_k}{c_k} \right)^{-1}$ 은 $Q \times Q$ 행렬, $\sum_{k=1}^n \frac{x_k y_k a_k}{c_k}$ 는 $Q \times 1$ 벡터를 나타낸다.

회귀추정량 \hat{t}_y 은

$$\hat{t}_y = \sum_{k \in s} y_k a_k g_k$$

와 같이 사후층화 가중치를 포함하는 식으로 재 표현 할 수 있다. 여기서

$$g_k = 1 + (t_x - \hat{t}_{xa})^t \left(\sum_{k \in s} \frac{x_k x_k^t a_k}{c_k} \right)^{-1} \frac{x_k}{c_k}$$

로써 일명 g-factor로 불리우는 사후층화 가중치이다. 가구 구성원에 대한 사후층화 가중치는

$$g_i = 1 + (t_x - \hat{t}_{xa})^t \left(\sum_{i \in s} z_i z_i^t a_i \right)^{-1} z_i$$

이며, 여기에서 $z_i = \frac{1}{c_k} \sum_{i=1}^q x_i$ 이다. 즉 사후층화 가중치의 특징은 가구에 대한 가중치와 가구 내의 구성원에 대한 가중치가 일치하여 모든 구성원들이 동일한 가중치를 갖는다는 점이다.

LFS의 최종가중치는 설계 가중치, 무응답 조정 가중치와 사후층화 가중치의 곱으로 표현되며, 이러한 최종가중치를 이용하여 주 지역 및 전국 단위의 경제활동인구 총계, 취업자 총계, 실업자 총계, 취업 및 실업률 등이 추정된다.

LFS에서 추정량의 분산계산은 잭나이프 방법을 이용한다. 좀 더 일반적인 경우의 잭나이프 방법에 대한 기술은 Wolter(1985)를 참조할 수 있으며, 여기에서는 LFS에서 적용하는 잭나이프 알고리즘을 소개하기로 한다.

(i) 잭나이프 방법을 적용하기 위해 h 번째 층은 J_h 개의 반복표본을 갖는다고 가정한다($a = 1, 2, \dots, J_h$). 우선 특정 반복표본에 해당하는 모든 가구들을 제거한다. 여기에서 해당 표본에서 반복 총계는 $J = \sum_{h=1}^H J_h$ 이고, H 는 해당 표본에서 층의 총계를 나타낸다.

(ii) 주어진 층에서 나머지 $J_h - 1$ 개의 반복표본의 모든 가구들에 대해 부차가중치에 대한 보정이 이루어진다. 보정된 가중치의 값은

$$a_k^{adj} = \frac{J_h}{J_h - 1} a_k \text{이다.}$$

(iii) 보정된 부차가중치와 남아있는 해당표본을 이용하여 관심 추정치 $\hat{t}_{y(ha)}$ 를 계산하기 위한 최중가중치를 계산한다. 여기에서 (ha)는 h번째 층으로부터 a번째 반복이 제거되었다는 것을 나타낸다.

해당표본의 모든 반복에 대해서 위의 (i)~(iii)의 절차가 반복되며, 결과로써 관심 추정치에 대한 J개의 서로 다른 추정값을 얻게 된다. 이러한 추정값을 이용하여 추정값의 분산을 계산하며 다음과 같은 분산 추정 공식을 이용한다.

$$\hat{V}(\hat{t}_y) = \sum_{h=1}^H \frac{J_h - 1}{J_h} \sum_{a=1}^{J_h} (\hat{t}_{y(ha)} - \hat{t}_y)^2.$$

실업률에 대한 분산 추정은 다음의 추정공식을 이용할 수 있다.

$$\hat{V}\left(100 \frac{\hat{t}_y}{\hat{t}_z}\right) = 100^2 \sum_{h=1}^H \frac{J_h - 1}{J_h} \sum_{a=1}^{J_h} \left(\frac{\hat{t}_{y(ha)}}{\hat{t}_{z(ha)}} - \frac{\hat{t}_y}{\hat{t}_z} \right)^2.$$

여기에서 y는 실업자 총계, z는 경제활동인구 총계를 나타내며, 위의 결과는 실업률 $100(y/z)\%$ 에 대한 잭나이프 분산 추정공식이다.

월 변화량의 추정치에 대한 잭나이프 분산추정을 다음과 같이 산출할 수 있다. 연속되는 두 달의 월 추정치로부터 다음의 차분추정치 $\hat{D}_y = \hat{t}_y^{(2)} - \hat{t}_y^{(1)}$ 을 고려하자. 여기에서 윗 첨자는 연속되는 월을 나타낸다.

대응되는 잭나이프 추정치는 $\hat{D}_{y(ha)} = \hat{t}_{y(ha)}^{(2)} - \hat{t}_{y(ha)}^{(1)}$ 으로 표현할 수 있다.

이때 분산 추정치는 다음과 같이 주어질 수 있다.

$$\hat{V}(\hat{D}_y) = \sum_{h=1}^H \frac{J_h - 1}{J_h} \sum_{a=1}^{J_h} (\hat{D}_{y(ha)} - \hat{D}_y)^2.$$

LFS의 연속된 두 달의 조사에서는 5/6의 표본이 일치한다. 공통표본을 이용하여 두 달 간의 월 변화량의 차를 추정하는 것이 위의 추정방법에 비해 훨씬 효율적일 수 있다. Singh et al.(1997)은 이러한 공통표본을 이용하여 다음과 같은 합성추정량을 제안하였다.

$$est_{(t+1)}^C = K \times est_{(t+1)} + (1-K) \times [est_{(t)}^C + change_{common}]$$

$$est_{(t+1)}^C = K \times est_{(t+1)} + (1-K) \times [est_{(t)}^C + change_{common}],$$

여기에서 윗첨자 C는 복합추정법을 의미하며, $change_{common}$ 은 공통표본을 이용한 변화량을 나타낸다. 이 추정량은 1998년 현재 캐나다 통계청에서 채택하고 있지는 않지만 새로운 표본설계에서는 사용될 전망이다.

위와 유사한 방법으로 n 개의 월 추정치의 평균에 대한 분산추정을 고려할 수 있다. n 개의 월 추정치의 평균은

$$\hat{A}_y = \sum_{i=1}^n \frac{\hat{t}_y^{(i)}}{n}$$

이고, 이에 대응하는 잭나이프 추정값은

$$\hat{A}_{y(ha)} = \sum_{i=1}^n \frac{\hat{t}_{y(ha)}^{(i)}}{n}$$

으로 계산할 수 있으며, 추정치의 분산은 다음의 추정공식을 이용할 수 있다.

$$\hat{V}(\hat{A}_y) = \sum_{h=1}^H \frac{J_h - 1}{J_h} \sum_{a=1}^{J_h} (\hat{A}_{y(ha)} - \hat{A}_y)^2.$$

3.2.6 데이터 관리

모든 표본조사에서와 마찬가지로 LFS 추정값들도 표본 오차와 비표본 오차를 수반한다. 따라서 조사 추정값들이 올바르게 해석되기 위해서는 추정값들의 정도를 나타내는 측도에 관한 점검이 요구된다.

표본오차에 직접적으로 영향을 미치는 것은 표본크기라 할 수 있다. 일반적으로 표본크기가 증가함에 따라 표본오차는 감소한다. 표본의 크기 뿐만 아니라 모집단의 변동, 추정 및 표본설계의 방법과 같은 요인들이 표본오차에 관련된 요인들이라 할 수 있다. 표본오차는 층화방법, 표본할당, 추출단위의 선택에서 뿐만 아니라 다 단계 표본설계에서 매 단계에서 선택된 표본 추출방법 및 추정방법 등과 같은 요인들에 크게 의존한

다.

표본설계 및 추정방법에 대한 효율성을 점검하기 위한 측도로는 평균 제곱오차 *MSE*가 이용된다. *MSE*는 추정값과 모집단의 실제값과의 편차 제곱합의 평균으로 보통 정의된다. 표본오차와 관련된 또 다른 중요한 측도로써 변동계수 *CV*가 이용된다. *Y*가 어떤 특성치에 대한 추정치이고, *d*가 추정치의 표준오차라 할 때, *CV*값은 $(d/Y) \times 100$ 으로 정의된다. 또한 추정치의 신뢰구간은 표준오차 *d*로부터 추론될 수 있다. 한편 시간에 따른 표본설계의 낙후성을 평가하기 위한 지표로써 설계효과(*design effect*)가 이용된다. 설계효과는 기존 표본설계로부터 조사된 추정값의 분산과 단순임의표본으로부터 계산된 추정값의 분산의 비로써 정의되며, LFS에서는 표본설계효과(*sample design effect*)와 전체설계효과(*overall design effect*)와 같은 두 가지 형태의 설계효과를 계산한다. 표본설계효과는 모총계에 대한 가중치의 조정 없이 부차가중치만을 이용하여 계산되며, 전체설계효과는 앞서 언급되었던 최종가중치를 반영하여 계산된다. 따라서 표본설계효과는 표본설계의 효율성만이 반영되며, 전체설계효과는 총화, 다단계추출, 사후총화 및 추정 등의 표본설계의 전반적인 사항들이 반영된다고 보면 된다.

1997년 1월부터 7월까지의 LFS 조사자료에 대한 주 단위 및 전국 단위의 고용 및 실업관련 추정치들의 *CV* 값이 다음 <표3.16>에 주어졌다.

<표3.16> LFS 고용 및 실업 추정치들의 CV 값 (1997)

Province	Employed CV(%)	Unemployed CV(%)	Province	Employed CV(%)	Unemployed CV(%)
Newfoundland	2.2	6.1	Manitoba	0.91	6.5
Prince Edward Island	1.7	6.5	Saskatchewan	1.1	7.4
Nova Scotia	1.2	5.3	Alberta	0.76	5.9
New Brunswick	1.2	5.5	British Columbia	0.90	5.1
Quebec	0.79	3.5	Canada	0.32	1.72
Ontario	0.54	3.0			

고용과 실업에 대한 월 변화 추정값의 표준오차는 <표3.17>에, 고용과 실업에 관한 설계효과는 <표3.18>에 주어졌다.

<표3.17> LFS 고용 및 실업에 대한 월변화 추정값의 표준오차(1997)
Unit: thousands

Province	SE (employed)	SE (Unemployed)	Province	SE (employed)	SE (Unemployed)
Newfoundland	3	2	Manitoba	4	3
Prince Edward Island	1	1	Saskatchewan	3	2
Nova Scotia	4	3	Alberta	9	6
New Brunswick	3	2	British Columbia	12	9
Quebec	18	14	Canada	32	24
Ontario	20	15			

<표3.18> LFS 고용 및 실업에 대한 설계효과 (1997)

Province	Employed		Unemployed	
	Sample	Overall	Sample	Overall
Newfoundland	2.7	0.83	1.4	1.3
Prince Edward Island	2.0	0.53	1.1	1.1
Nova Scotia	2.2	0.51	1.2	1.1
New Brunswick	2.0	0.56	1.4	1.4
Quebec	2.1	0.55	1.1	1.0
Ontario	3.3	0.50	1.2	1.1
Manitoba	2.2	0.41	1.1	1.1
Saskatchewan	2.4	0.63	1.2	1.2
Alberta	4.1	0.40	1.1	1.1
British Columbia	2.1	0.50	1.2	1.1
Canada	2.8	0.51	1.2	1.1

비표본오차는 표본조사의 매 단계에서 발생할 수 있으며 주로 조사원의 무관심, 오해 및 잘못된 해석 등의 사유에 기인하며, 추정값의 편향 및 변동에 직접적인 영향을 미친다. 관측값의 수가 많거나 혹은 대영역의 조사에서는 비표본오차에 기인한 효과는 무시될 수도 있는 양이나, 소지역 추정치 문제에서는 민감한 문제로 인식되어진다. 비표본 편향 및 분산은 조사원에 대한 교육 및 조사원의 태도, 설문지 설계 상의 문제 또는 무응답을 처리하기 위해 이용되는 대체방법 등에서 발생할 수 있으며, 여기에서는 LFS에서의 적용범위 오차(coverage error), 무응답 오차, 빈집 오차(vacancy error), 응답 오차(response error), 처리과정 오차(processing error)등에 관해서 설명하기로 한다.

적용범위 오차는 표본추출틀의 조사단위들이 목표모집단을 제대로 반영하지 못할 경우 발생할 수 있다. 조사단위들이 표본추출틀에서 누락되어 있는 경우, 목표모집단에 속하지 않은 단위들이 표본추출틀에 포함되어 있는 경우 또는 조사단위들이 표본추출틀에 중복되어 있는 경우 등이 적용범위 오차를 유발할 수 있는 일반적인 유형들이며, 이 중 조사단위들이 표본추출틀에서 누락되어 있는 경우가 LFS에서 가장 빈번히 발생하는 유형이라 할 수 있다. 나머지 유형의 문제는 LFS에서는 거의 무시된다. LFS에서는 적용범위 오차를 측정하기 위한 지표로써 손실률(slippage rate)을 이용한다. 손실률은 LFS 인구 추정치와 최근의 센서스 인구 추정치와의 차이에 대한 센서스 인구 추정치의 비율로써 정의된다. LFS에서는 CMA 지역, ER 지역, 주 및 전국 단위와 캐나다 지역의 성별(남, 녀)-연령대별(15-19, 20-24, 25-29, 30-39, 40-54, 55+) 범주에 대한 손실률을 매월 정기적으로 작성하고 있다. 손실률로부터 발생하는 비표본오차에 대한 보정은 추정과정에서 처리하고 있다. LFS 조사에서 평균적인 손실률의 양은 다음 <표3.19>와 같이 주어진다.

<표3.19> LFS 평균 손실률(Average Slippage Rate(%))

Provinces		Average	Provinces	Average
Canada	all	9.3	Nova Scotia	8.6
	15-19세	6.1	New Brunswick	10.4
	20-24세	15.6	Quebec	8.0
	25-29세	16.1	Ontario	9.7
	30-39세	9.8	Manitoba	6.1
	40-54세	8.0	Saskatchewan	10.7
	55이상	6.9	Alberta	7.4
Newfoundland		9.8	British Columbia	12.4
Prince Edward Island		11.6		

조사가구에 대한 무응답 발생요인으로써 가구 구성원의 부재, 가구 구성원의 비 정상적인 거주 환경, 인터뷰 거절 등의 요인을 들 수 있다. 여기에서 인터뷰 거절의 비율은 매월 조사에서 1~2% 정도로 매우 낮게 나타나며, 주 지역에 대해서도 월 조사와 비슷한 비율을 보이거나 높게는 약 3% 정도까지 나타난다. 단위 무응답에 대해서는 바로 전 달의 정보를 이용할 수 있다면 이를 이용하여 대체되며, 항목 무응답에 대해서는 표본 대체법이 이용된다. 가구 구성원이 거주하고 있지 않는 결측 가구 및 건물 철거 등으로 인한 비 존재 가구들에 대해서 발생하는 무응답은 편향에 영향을 미치지 않는지만 표본 분산에는 영향을 미치게 되므로 LFS에서는 이러한 유형의 오차 정보를 파악하기 위한 VC(vacancy check) 프로그램을 운영하고 있다. 다음 <표3.20>은 1997년 LFS 조사에서 발생한 평균 무응답률을 나타낸다.

<표3.20> LFS 평균 무응답률(1997)

Provinces	Average(%)	Maximum(%)	Minimum(%)
Newfoundland	4.2	5.4	3.0
Prince Edward Island	3.5	4.8	2.4
Nova Scotia	6.3	7.3	4.6
New Brunswick	4.6	5.4	3.1
Quebec	5.4	6.6	3.7
Ontario	4.8	5.7	3.7
Manitoba	3.6	5.4	2.1
Saskatchewan	3.6	4.6	2.4
Alberta	4.9	6.3	3.1
British Columbia	5.7	6.7	4.5
Canada	4.9	5.5	3.8

LFS에서 결측가구(dwelling vacant)는 사람이 거주하고 있지 않는 가구, 계절 가구 또는 공사 중인 가구로 정의되어 분류된다. 철거 또는 조사 가구가 상점 등으로 용도가 변경된 경우는 비 존재 가구(dwelling non-existence)로 분류된다. 결측가구로 확인된 가구들은 LFS 추정치의 편향에 영향을 미치지 않는 않지만 표본 조사단위가 줄어들므로 추정 분산은 커지게 된다. 결측가구들은 새로운 입주자들이 상주할 가능성이 항상 존재하므로 매월 조사 대상에 포함된다. 그러나 LFS 조사에서 비 존재 가구로 확인된 조사가구들은 표본추출틀에서 일제히 삭제된다. 1997년 LFS 조사에서 발생한 평균적인 결측가구에 대한 비율이 다음 <표3.21>에 주어졌다.

<표3.21> LFS 평균 결측률(vacant rate)(1997)

Provinces	Average(%)	Maximum(%)	Minimum(%)
Newfoundland	15.4	14.9	16.4
Prince Edward Island	20.5	18.6	23.0
Nova Scotia	16.8	15.2	18.7
New Brunswick	14.1	13.5	15.2
Quebec	14.0	11.9	15.8
Ontario	10.8	10.0	11.3
Manitoba	17.1	16.4	17.7
Saskatchewan	14.7	12.5	15.5
Alberta	8.7	8.1	9.8
British Columbia	9.5	8.7	9.8
Canada	13.0	12.2	13.5

응답 오차(response error)는 설문지 설계, 문항 구성, 응답자의 인지력, 인터뷰 방식, 조사가 수행되는 상황 및 조사정보가 수집되고 집계되는 과정 등에 기인할 수 있다. 조사정보가 수집되고 집계되는 과정에서 발생하는 응답오차는 CAI 시스템에 의해 어느 정도 보완되었다고 볼 수 있다.

처리과정 오차(processing error)는 자료 획득, 편집, 코딩, 가중치를 산출하는 과정 및 목록화 작업 등의 매 단계에서 발생할 수 있다. LFS 조사에서는 이러한 매 단계의 처리과정을 전산화 작업으로 통합하여 자료 처리과정에서 발생하는 오차를 최소화하고 있다. 전산화 통합 모드는 1993년부터 채택되어 시행되고 있으며 앞서 언급되었던 CAI 모드는 조사 행정에서 조사원들의 조사과정을 보조하는 일종의 컴퓨터 보조관리 시스템을 말한다.

3.2.7 소지역 추정법

캐나다 노동력 조사에서는 표본설계 단계에서 EIER 지역과 CMA 지역과 같은 소지역 단위에 대한 추정을 고려하여 층화, 표본추출, 표본 배정 등이 이루어진다. 소지역 통계 추정을 위한 추정량은 크게 설계 기반 추정량(design-based estimator), 간접추정량(indirect estimator), 모형 기반 추정량(model-based estimator)이 이용된다. 소지역 통계 작성 시 설계 기반 추정량이 목표 요구정도를 만족한다면 우선적으로 설계 기반 추정량을 이용하며 그렇지 못할 경우에는 추정량의 신뢰도를 확보할 수 있는 다른 추정방법을 이용한다.

(1) 설계 기반 추정량(Design-Based Estimator)

일반적으로 설계 기반 추정량은 직접추정량(direct estimator)과 수정된 직접추정량(modified direct estimator)으로 구분된다. 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 추정되기도 한다. 가장 간단한 총계추정에 대한 직접추정량으로써 다음과 같은 단순추정량(expansion estimator)이 이용된다.

$$\hat{Y}_{e,a} = \sum_{i \in s_a} w_i y_i \quad (3.18)$$

여기에서 s_a 는 소지역 a 의 표본들의 집합, w_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 알고 있을 경우에는 다음과 같은 사후층화추정량이 이용된다.

$$\begin{aligned} \hat{Y}_{pst,a} &= N_a \frac{\sum_{i \in s_a} w_i y_i}{\sum_{i \in s_a} w_i} \\ &= N_a \frac{\hat{Y}_{e,a}}{\hat{N}_{e,a}} \\ &= N_a \bar{y}_{e,a} \end{aligned} \quad (3.19)$$

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 이용된다.

$$\begin{aligned} \hat{Y}_{st,pst,a} &= \sum_h \left(N_{h,a} \frac{\sum_{i \in s_{h,a}} w_i y_i}{\sum_{i \in s_{h,a}} w_i} \right) \\ &= \sum_h N_{h,a} \frac{\hat{Y}_{h,e,a}}{\hat{N}_{h,e,a}} \\ &= \sum_h N_{h,a} \bar{y}_{h,a} \end{aligned} \quad (3.20)$$

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.

$$\hat{Y}_{r,a} = X_a \hat{R}_a, \quad \hat{Y}_{st,r,a} = \sum_h X_{h,a} \hat{R}_{h,a} \quad (3.21)$$

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.

$$\hat{Y}_{reg,a} = \hat{Y}_a + \hat{\beta}_a (X_a - \hat{X}_a) \quad (3.22)$$

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 와 유사한 방법으로 추정된다. 추정모수 $\hat{\beta}_a$ 는 관심변수 y 와 공변량 x 의 관계로부터 추정되며

$$\hat{\beta}_a = \sum_{i \in s_a} v_i^{-1} w_i y_i x_i^t \left(\sum_{i \in s_a} v_i^{-1} w_i x_i x_i^t \right)^{-1}$$

과 같이 주어진다. 여기에서 v_i 는 회귀가중치로써 주어지는 값이며, $v_i = x_i$ 이고 x 가 상수일 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 불편성

은 회귀추정량과 마찬가지로 근사적으로 만족된다. 예를 들면 식(3.22)에서 추정 회귀모수 $\hat{\beta}_a$ 대신에 회귀모수에 대한 합성추정량의 일종인

$$\hat{\beta} = \sum_{i \in s} v_i^{-1} w_i y_i x_i^t \left(\sum_{i \in s} v_i^{-1} w_i x_i x_i^t \right)^{-1}$$

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 $v_i = x_i$ 인 경우에는 $\hat{\beta}$ 대신 $\hat{R} = \hat{Y}_o / \hat{X}_o$ 이 이용될 수도 있다.

(2) 간접추정량(Indirect Estimator)

간접추정량은 합성추정량(synthetic estimator), 복합추정량(composite estimator), 표본수 의존 복합추정량(sample size dependent estimator) 등의 유형으로 구분되며, 해당 지역의 조사자료 뿐만 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 특성구조가 유사하다는 가정 하에서 이용된다. 합성추정량의 분산은 직접추정량의 분산에 비해 작으나 전체한 가정이 성립하지 않을 경우에는 심각한 편향이 발생할 수 있다.

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

$$\hat{Y}_{syn, m, a} = N_a \frac{\sum_{i \in s} w_i y_i}{\sum_{i \in s} w_i} = N_a \bar{y} \quad (3.23)$$

층화 또는 사후층화에 근거한 합성추정량은 보통 다음과 같은 형태로 주어진다.

$$\hat{Y}_{syn, st, m, a} = \sum_h N_{h, a} \frac{\sum_{i \in s_h} w_i y_i}{\sum_{i \in s_h} w_i} = \sum_h N_{h, a} \bar{y}_h \quad (3.24)$$

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

$$\hat{Y}_{syn, r, a} = X_a \frac{\hat{Y}_e}{\hat{X}_e}, \quad \hat{Y}_{syn, st, r, a} = \sum_h X_{h, a} \frac{\hat{Y}_{h, e}}{\hat{X}_{h, e}} \quad (3.25)$$

여기에서 $\hat{Y}_e = \sum_{i \in s} w_i y_i$ 는 y 에 대한 모집단 총계 추정량이고,

$\hat{Y}_{h, e} = \sum_{i \in s_h} w_i y_i$ 는 층의 총계 추정치를 나타낸다. 기타 비합성추정량들에

대한 내용은 Gonzalez(1973), Gonzalez and Waksberg(1973), Ghangurde and Singh(1977, 1978)에 자세히 소개되어있다. 한편, Singh and

Tessier(1976)는 (3.25)식의 $\hat{Y}_{syn, r, a}$ 에서 \hat{X}_e 대신에 X 를 이용한 비합성추

정량의 대체식 $\bar{Y}_{syn, r, a} = X_a \frac{\hat{Y}_e}{X}$ 를 제안하였다. 여기에서 $\hat{Y}_{syn, r, a}$ 와

$\bar{Y}_{syn, r, a}$ 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 표

본의 크기가 클 경우, 두 추정량의 분산은 상관계수의 값이 $\rho \geq 0.5c_x/c_y$

일 때 $V(\hat{Y}_{syn, r, a}) \leq V(\bar{Y}_{syn, r, a})$ 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. 상관 계수 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 $\bar{Y}_{syn, r, a}$ 를 선택한다.

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

$$\hat{Y}_{syn, r, a}^{(2)} = \gamma_a X_a \frac{\hat{Y}_e}{\hat{X}_e} + (1 - \gamma_a) Z_a \frac{\hat{Y}_e}{\hat{Z}_e} \quad (3.26)$$

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

회귀합성추정법은 비합성추정법과 유사하며 추정량은 다음과 같이 주어진다.

$$\hat{Y}_{syn, reg, a} = \hat{\beta} X_a, \quad \hat{\beta} = \sum_{i \in s} v_i^{-1} w_i y_i x_i^t \left(\sum_{i \in s} v_i^{-1} w_i x_i x_i^t \right)^{-1} \quad (3.27)$$

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

$$\hat{Y}_{syn, Roy, a} = \sum_{i \in s_a} y_i + \hat{\beta} (X_a - \sum_{i \in s_a} x_i) \quad (3.28)$$

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

$$\hat{Y}_{com, a} = \lambda_a \hat{Y}_{dir, a} + (1 - \lambda_a) \hat{Y}_{syn, a} \quad (3.29)$$

여기에서 가중치 λ_a 는 적절히 선택되는 값이다.

가중치 λ_a 를 결정하는 방법은 크게 세 가지 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 표본 크기에 의존하는 복합추정량으로써 다음과 같은 추정량을 제안하였다.

$$\hat{Y}_{ssd,r,a} = \lambda_a \hat{Y}_{r,a} + (1 - \lambda_a) \hat{Y}_{syn,r,a} \quad (3.30)$$

여기에서

$$\lambda_a = \begin{cases} 1 & \text{if } \hat{N}_{e,a} \geq \delta N_a \\ \frac{\hat{N}_{e,a}}{\delta N_a} & \text{otherwise} \end{cases}$$

이며, δ 는 합성추정량 부분의 편향을 보정하기 위해 주관적으로 결정되는 값이다. 캐나다 노동력조사에서는 $\delta = 2/3$ 를 이용한다. 식 (3.30)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 노동력조사에서 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

$$\hat{Y}_{ssd,reg,a} = \lambda_a \hat{Y}_{sreg,a} + (1 - \lambda_a) \hat{Y}_{syn,reg,a}$$

를 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 Hidirolou(1989)는 Rao의 가중치에서 $\hat{N}_{e,a} < N_a$ 일 때 $\lambda_a = (\hat{N}_{e,a}/N_a)^{h-1}$ 을 사용할 것을 제안하였다. 여기에서 h 는 합성추정량의 편향을 감안하여 적절히 선택되는 값이다. 가중치를 결정하는 세 번째 방법은 직접추정량과 합성추정량의 평균제곱오차와 두 추정량의 공분산을 자료로부터 추정하여 최적가중치를 산정하는 방법이다. 복합추정량의 평균제곱오차는 다음 식과 같이 나타낼 수 있다.

$$MSE(\hat{Y}_{com,a}) = \lambda_a^2 MSE(\hat{Y}_{dir,a}) + (1 - \lambda_a)^2 MSE(\hat{Y}_{syn,a}) + 2\lambda_a(1 - \lambda_a)E(\hat{Y}_{dir,a} - Y_a)(\hat{Y}_{syn,a} - Y_a) \quad (3.31)$$

(3.31)식에서 $\hat{Y}_{com,a}$ 의 MSE 를 최소화하는 가중치 λ_a 는 다음 식과 같이 주어질 수 있다.

$$\hat{\lambda}_a = \frac{\widehat{MSE}(\hat{Y}_{syn,a}) - E(\hat{Y}_{syn,a} - Y_a)(\hat{Y}_{dir,a} - Y_a)}{\widehat{MSE}(\hat{Y}_{syn,a}) + \widehat{MSE}(\hat{Y}_{dir,a}) - 2E(\hat{Y}_{syn,a} - Y_a)(\hat{Y}_{dir,a} - Y_a)} \quad (3.32)$$

식 (3.32)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 수도 있다.

$$\hat{\lambda}_a^* = \frac{\widehat{MSE}(\hat{Y}_{syn,a})}{\widehat{MSE}(\hat{Y}_{syn,a}) + \widehat{MSE}(\hat{Y}_{dir,a})} \quad (3.33)$$

(3) 모형 기반 추정량(Model-Based Estimator)

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 시

계열자료(time series data)를 함께 추정과정에 이용하는 방법에 관한 연구가 활발히 진행되고 있다. 이 절에서는 횡단면 조사자료를 이용한 모형 기반 추정량과 횡단면 조사자료와 시계열자료를 함께 이용하는 모형기반 추정량에 대한 내용들을 소개한다.

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 가정할 때, 소지역 i 에 대한 다음과 같은 선형회귀모형(linear regression model)을 고려할 수 있다.

$$\theta_i = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, I \quad (3.34)$$

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

$$\bar{\theta}_{i(\text{reg})} = \bar{\beta}_0 + \bar{\beta}_1 x_i \quad (3.35)$$

여기에서 $\bar{\beta}_0$ 와 $\bar{\beta}_1$ 은 결합모형 $y_i = \beta_0 + \beta_1 x_i + e_i$ 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 추정량이나 $EBLUP$ 추정량은 적당한 가중치가 부여되어 편향 발생이 다소 억제되는 결과를 얻을 수 있다.

이러한 편향에 대한 문제점을 해결하기 위해 Fay and Herriot(1979)는 모형 (3.34)를 다음과 같이 해당 소지역에 대한 랜덤효과 v_i 를 갖는 모형으로 보완하였다.

$$\theta_i = \beta_0 + \beta_1 x_i + v_i \quad (3.36)$$

여기에서 v_i 는 평균이 0 이고 분산이 σ_v^2 을 갖는 서로 독립인 정규분포를 따르는 확률변수, e_i 는 평균이 0 이고 분산이 σ_e^2 인 서로 독립인 정규 확률변수를 나타내며, σ_e^2 은 기지인 값으로 가정된다. 이때 결합모형은 다음

과 주어진다.

$$y_i = \beta_0 + \beta_1 x_i + v_i + e_i \quad (3.37)$$

위의 모형으로부터 θ_i 의 EB 추정량은 직접조사추정량 y_i 와 회귀합성 추정량 $\hat{\theta}_{i(reg)} = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 의 가중합으로 표현되며 다음 식과 같이 주어진다.

$$t_i(\hat{\sigma}_v^2, y) = w_i y_i + (1 - w_i) \hat{\theta}_{i(reg)} \quad (3.38)$$

여기에서 $w_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \sigma_i^2)$, $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 은 결합모형 하에서 추정된 가중 최소제곱추정량을 나타내며, $\hat{\sigma}_v^2$ 은 σ_v^2 의 적률추정량 또는 최대우도추정량 등이 이용될 수 있다. Fay and Herriot(1979)는 1970년 미국의 인구주택 총조사 자료로부터 인구 1000 미만의 소지역에 대한 소득관련 추정에 식 (3.38)의 EB추정량을 이용하였고, EB추정량이 직접 조사 추정량이나 합성추정량에 비해 표본오차가 작다는 사실을 수치적으로 제시하였다.

횡단면자료를 이용한 소지역 추정방법은 조사 시기가 상이한 조사자료들의 정보를 모형에 반영시키는 것은 사실상 어렵다. Scott et al.(1977), Jones(1980), Tiller(1989) 등은 이러한 단점을 보완하기 위해 반복적인 월별 조사자료들의 정보와 센서스 및 행정자료를 모형에 포함시킨 횡단면 시계열 모형들을 소지역 추정 문제에 도입하였다.

θ_{it} , y_{it} 와 x_{it} 를 각각 조사시기 t 에서 소지역 i 에 대한 모평균, 직접조사추정값, i 번째 소지역과 관계가 있는 연관변수라 할 때, 우선 다음과 같은 모형을 고려한다.

$$y_{it} = \theta_{it} + e_{it}, \quad i = 1, 2, \dots, I; t = 1, 2, \dots, T \quad (3.39)$$

여기에서 표본오차 e_{it} 의 평균은 0, 분산공분산행렬은 기지인 블록대각행렬 Σ_i ($T \times T$ 행렬)로 가정한다. 모평균 θ_{it} 에 관한 모형은 다양한 유형으로 설정될 수 있으며 다음과 같은 모형들이 고려될 수 있다.

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

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

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

여기에서 $v_i \sim N(0, \sigma_v^2)$, $\epsilon_{it} \sim N(0, \sigma^2)$, $\{v_i\}$ 와 $\{\epsilon_{it}\}$ 는 서로 독립이며,

모형(I)과는 달리 v_i 들이 랜덤효과로 가정되었다.

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

여기에서 $v_i \sim N(0, \sigma_v^2)$, $u_{it} \sim N(0, \sigma_u^2)$, $\epsilon_{it} \sim N(0, \sigma^2)$ 이고, $\{v_i\}$, $\{u_{it}\}$

와 $\{\epsilon_{it}\}$ 는 서로 독립이다. v_i 는 소지역에 대한 랜덤효과, u_{it} 는 조사시기에 대한 랜덤효과이다.

$$(IV) \theta_{it} = \beta_0 + \beta_1 x_{it} + v_i + u_{it}, \quad u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1,$$

여기에서 $v_i \sim N(0, \sigma_v^2)$, $\epsilon_{it} \sim N(0, \sigma^2)$ 이고, $\{v_i\}$ 와 $\{\epsilon_{it}\}$ 는 서로 독립이

며, $\{u_{it}\}$ 는 AR(1)과정을 따른다. 모형 (IV)는 다음과 같이 시차모형(lag model)으로 재표현 가능하다.

$$\theta_{it} = \rho \theta_{i,t-1} + (1-\rho)\beta_0 + \beta_1 x_{it} - \beta_1 \rho x_{i,t-1} + (1-\rho)v_i + \epsilon_{it} \quad (3.40)$$

모형 (IV)는 모평균 θ_{it} 와 보조변수 x_{it} 가 이전 조사시기에서의 값들을 모형에 반영시키기 때문에 위의 네가지 모형들 중에서는 가장 현실적인 추정 모형이라 할 수 있다. 따라서 모형 (IV)를 이용한 결합모형을 고려한다면 다음과 같이 주어진다.

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_i + (e_{it} + u_{it}) \quad (3.41)$$

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1$$

여기에서 $v_i \sim \overset{ind}{N}(0, \sigma_v^2)$, $\epsilon_{it} \sim \overset{ind}{N}(0, \sigma^2)$, e_{it} 는 평균이 0 이고, 가지인 불
 력대각공분산행렬 $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_I)$ 를 갖는다. 캐나다 노동력조
 사에서는 표본 공분산행렬 Σ 를 이용할 수 없기 때문에 Tiller(1989)는 복
 합오차 $w_{it} = e_{it} + u_{it}$ 를 $AR(1)$ 과정으로 취급한 후 다음과 같은 결합모형
 을 이용하여 소지역 추정을 시도하였다.

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_i + w_{it}, \quad (3.42)$$

$$w_{it} = \rho w_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1$$

여기에서 $\theta_{it} = \beta_0 + \beta_1 x_{it} + v_i$, $v_i \sim \overset{ind}{N}(0, \sigma_v^2)$, $\epsilon_{it} \sim \overset{ind}{N}(0, \sigma^2)$.

$\{y_{it}\}$ 를

$$\begin{aligned} y &= (y_{11}, y_{12}, \dots, y_{1T}; \dots; y_{I1}, y_{I2}, \dots, y_{IT})^t \\ &= (y_1^t, y_2^t, \dots, y_I^t)^t \end{aligned}$$

로 표현하면 위의 모형 (3.42)는 다음과 같은 일반적인 혼합모형(mixed model)의 일종으로 볼 수 있다.

$$y = X\beta + Zv + w, \quad v \sim (0, \sigma_v^2 I), \quad w \sim (0, \sigma^2 (I \otimes \Gamma)) \quad (3.43)$$

여기에서

$$X^t = (X_1^t, X_2^t, \dots, X_I^t),$$

$$Z = I \otimes 1_T,$$

$$\beta = (\beta_0, \beta_1)^t$$

이고 X_i^t 는 t 번째 행이 $(1, x_{it})$ 로 주어지는 $T \times 2$ 행렬, I 는 크기 I 인 항등
 행렬, 1_T 는 1을 원소로 갖는 t -벡터, Γ 는 (i, j) 번째 원소가

$$\gamma_{ij} = (1 - \rho^2)^{-1} \rho^{|i-j|}$$

인 $T \times T$ 행렬을 나타낸다.

β 와 v 의 일차결합 $\tau = k^t\beta + m^tv$ 에 대한 $\bar{\tau} (= \bar{\theta}_{ii})$ 의 *BLUP*와 *BLUP*의 *MSE*는 다음과 같이 주어진다(Henderson, 1975).

$$\bar{\tau} = k^t\bar{\beta} + m^tZ^t\Sigma^{-1}(y - X\bar{\beta})(\sigma_v^2/\sigma^2) \quad (3.44)$$

$$\begin{aligned} MSE(\bar{\theta}_{ii}) = & \sigma^2 \left\{ k^t(X^t\Sigma^{-1}X)^{-1}k + \frac{\sigma_v^2}{\sigma^2}m^tm - \left(\frac{\sigma_v^2}{\sigma^2} \right)^2 m^tZ^t\Sigma^{-1}AZm \right. \\ & \left. - 2\frac{\sigma_v^2}{\sigma^2}k^t(X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}Zm \right\} \end{aligned} \quad (3.45)$$

여기에서

$$\Sigma = I \otimes \{ (\sigma_v^2/\sigma^2)J + \Gamma \},$$

$$\bar{\beta} = (X^t\Sigma^{-1}X)^{-1}(X^t\Sigma^{-1}y),$$

$$A = I - X(X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}$$

이고, J 는 원소들이 1로 구성된 $T \times T$ 행렬을 나타낸다.

식 (3.44)를 이용하여 결합모형 (3.42)의 $\theta_{ii}(= \tau)$ 에 대한 *BLUP*추정량을 계산하면 주요 항들은 다음과 같이 주어질 수 있다.

$$k^t = (1, x_{ii}),$$

$$m^t = (0, \dots, 0, 1, 0, \dots, 0)_i,$$

$$m^tZ^t\Sigma^{-1}(y - X\bar{\beta}) = 1_T^t \{ (\sigma_v^2/\sigma^2)J + \Gamma \}^{-1} (y_i - X_i\bar{\beta})$$

식 (3.44)의 *BLUP*는 미지인 분산비(variance ratio) σ_v^2/σ^2 과 자기상관계수(autocorrelation) ρ 에 의존하므로 이들 값들은 추정되어야 한다. 모수들에 대한 *BLUP*는 다음과 같이 주어진다(Pantula and Pollack, 1985).

$$\hat{\rho} = \left\{ \sum_{i=1}^I \sum_{t=1}^{T-2} \hat{e}_{it} (\hat{e}_{i,t+1} - \hat{e}_{i,t+2}) \right\} \left\{ \sum_{i=1}^I \sum_{t=1}^{T-2} \hat{e}_{it} (\hat{e}_{it} - \hat{e}_{i,t+1}) \right\}^{-1} \quad (3.46)$$

$\hat{\sigma}_v^2$ 과 $\hat{\sigma}^2$ 은 다음 (3.47), (3.48), (3.49), (3.50)식의 정의로부터 유도될 수

있다.

$$z_{it}^{(1)} = z_{it} - z_{it}^{(2)}. \quad (3.47)$$

여기에서

$$z_{it} = \begin{cases} y_{it} - \hat{\rho}y_{i,t-1} & , t \geq 2 \\ f_1 y_{it} & , t = 1 \end{cases},$$

$$z_{it}^{(2)} = c^{-1} d_i f_t,$$

$$c = (1 - \hat{\rho})\{T - (T-2)\hat{\rho}\},$$

$$f_t = \begin{cases} 1 - \hat{\rho}^2 & , t = 1 \\ 1 - \hat{\rho} & , t \geq 2 \end{cases},$$

$$d_i = \sum_{t=1}^T f_t z_{it}$$

로 주어진다.

$$h_{0it}^{(1)} = h_{0it} - h_{0it}^{(2)}. \quad (3.48)$$

여기에서

$$h_{0it} = \begin{cases} 1 - \hat{\rho} & , t \geq 2 \\ f_1 & , t = 1 \end{cases},$$

$$h_{0it}^{(2)} = c^{-1} d_i f_t,$$

$$f_t = \begin{cases} 1 - \hat{\rho}^2 & , t = 1 \\ 1 - \hat{\rho} & , t \geq 2 \end{cases},$$

$$d_i = \sum_{t=1}^T f_t h_{0it}$$

로 주어진다.

$$h_{1it}^{(1)} = h_{1it} - h_{1it}^{(2)}. \quad (3.49)$$

여기에서

$$h_{1it} = \begin{cases} x_{it} - \hat{\rho}x_{i,t-1} & , t \geq 2 \\ f_1 x_{it} & , t = 1 \end{cases},$$

$$h_{1i}^{(2)} = c^{-1} d_i f_t,$$

$$f_t = \begin{cases} 1 - \hat{\rho}^2, & t = 1 \\ 1 - \hat{\rho}, & t \geq 2 \end{cases},$$

$$d_i = \sum_{t=1}^T f_t h_{1it}$$

로 주어진다.

$$g_i = \sum_{t=1}^T f_t z_{it}, \quad f_{0i} = \sum_{t=1}^T f_t h_{0it}, \quad f_{1i} = \sum_{t=1}^T f_t h_{1it} \quad (3.50)$$

종속변수를 $z_{it}^{(1)}$, 설명변수를 $h_{0it}^{(1)}$ 와 $h_{1it}^{(1)}$ 를 갖는 절편항이 없는 회귀식을 적합시켰을 때 얻어지는 잔차제곱합을 $\hat{e}^t e$ 라 하고, 종속변수 g_i , 설명변수 f_{0i} , f_{1i} 를 갖는 절편항이 없는 회귀식에서 얻어지는 잔차제곱합을 $\hat{u}^t \hat{u}$ 라 할때, σ_v^2 과 σ^2 의 *BLUP*는 최종적으로 다음과 같이 주어진다.

$$\hat{\sigma}_v^2 = c^{-1} (I-2)^{-1} \{ \hat{u}^t \hat{u} - \hat{\sigma}^2 (I-2) \}, \quad (3.51)$$

$$\hat{\sigma}^2 = \{ I(T-1) - 2 \}^{-1} \hat{e}^t \hat{e}$$

따라서 θ_{it} 의 *EBLUP* $\hat{\theta}_{it}$ 는 식 (3.44)에서 $\hat{\rho}$, $\hat{\sigma}_v^2$, $\hat{\sigma}^2$ 을 대체하여 얻을 수 있다.

만약 $p-1 (\geq 2)$ 개의 x 변수들이 모형에 포함되어 있다면, θ_{it} 의 *EBLUP* $\hat{\theta}_{it}$ 는 다음과 같이 추정하면 된다. 우선 y_{it} 와 $x_{1it}, \dots, x_{p-1, it}$ 의 회귀식으로부터 $\{\hat{e}_{it}\}$ 를 추정한다.

다음으로 $\{h_{j\bar{i}t}, h_{j\bar{i}t}^{(1)}, h_{j\bar{i}t}^{(2)}; j=0, 1, \dots, p-1\}$ 을 앞서 언급된 바와 같이 1, $x_{1it}, \dots, x_{p-1, it}$ 의 원소들로 정의한 후, $z_{it}^{(1)}$ 과 $h_{0it}^{(1)}, h_{1it}^{(1)}, \dots, h_{p-1, it}^{(1)}$ 의 절편항이 없는 회귀식으로부터 $\hat{e}^t e$ 를 추정한다. 같은 방법으로 $f_{\bar{j}} = \sum_{t=1}^T f_t h_{\bar{j}it}$ ($j=0, 1, \dots, p-1$)를 계산한 후, g_i 와 $f_{0i}, f_{1i}, \dots, f_{p-1, i}$ 의

절편항이 없는 회귀식을 적합시켜 $\hat{u}'\hat{u}$ 를 구한다. 마지막으로 (3.51)식에서 $I(T-1)-2$ 를 $I(T-1)-p$ 로, $I-2$ 를 $I-p$ 로 대체하여 $BLUP \hat{\sigma}_v^2$, $\hat{\sigma}^2$ 과 $\hat{\rho}$ 를 계산하면 $EBLUP \hat{\theta}_u$ 를 얻을 수 있고, (3.45)식으로부터 $EBLUP \hat{\theta}_u$ 의 MSE 값을 계산할 수 있다. 한편, 모형 (3.42) 하에서 조사 추정량 y_u 의 MSE 는 다음과 같이 주어진다.

$$MSE(y_u) = E(y_u - \theta_u)^2 = V(w_u) = \frac{\sigma^2}{1 - \rho^2} \quad (3.52)$$

$MSE(y_u)$ 의 추정량은 위의 (3.52)식에서 σ^2 , ρ 를 각각 $\hat{\sigma}^2$, $\hat{\rho}$ 로 대체하여 얻을 수 있다.

(4) 소지역 추정량들의 효율

$\hat{Y}_{M,a}(r)$ 을 소지역 추정법 M 을 이용하여 추정된 r 번째 반복에서 특성치 Y_a 의 몬테카를로 추정값이라 하자. 이 때 n 개의 소지역에 대한 MSE 추정값의 평균은 다음 식을 이용하여 계산할 수 있다.

$$Avg \widehat{MSE}_M = \frac{1}{n} \sum_a \sum_{r=1}^R \frac{(\hat{Y}_{M,a}(r) - Y_a)^2}{R} \quad (3.53)$$

소지역 추정법 M 을 이용하여 추정된 추정량들의 효율을 직접추정법 M_0 를 이용한 추정량과 비교하여 나타낸다면 상대 효율은 다음 식을 이용하여 구할 수 있다.

$$Eff(M \text{ vs } M_0) = \frac{Avg \widehat{MSE}_{M_0}}{Avg \widehat{MSE}_M} \quad (3.54)$$

여기에서 $Avg \widehat{MSE}_{M_0}$ 는 직접추정법 M_0 에 의해 추정된 추정값들에 대한 평균제곱오차의 평균을 나타낸다.

(5) 표본 수 의존 추정(sample size dependent estimation)

현재 캐나다에서 소지역 통계 작성을 위해 사용되고 있는 추정량은 Drew et al.(1982)에 기초한 표본수 의존 추정량(sample size dependent estimator)이다. LFS자료에 근거하여 우선 소지역 a 에 대해서 일반화 회귀추정량으로 관심변수의 총계를 추정한다.

$$\hat{Y}_{GREG,a} = \hat{Y}_{e,a} + \hat{\beta}_a (X_a - \hat{X}_{e,a}) \quad (3.55)$$

여기에서

$$\hat{Y}_{e,a} = \sum_{i \in s_a} w_i y_i,$$

$$\hat{X}_{e,a} = \sum_{i \in s_a} w_i x_i,$$

$$\hat{\beta}_a = \sum_{i \in s_a} w_i y_i x_i^t \left(\sum_{i \in s_a} w_i x_i x_i^t \right)^{-1}$$

이다.

관심변수의 총계를 추정할 수 있으면 관심의 대상인 경제활동인구수, 실업자 수, 취업자 수, 실업률, 취업률 등을 모두 추정할 수 있게 된다. 추정과정에 이용된 가중치 w_i 는 설계 가중치와 무응답 가중치 조정이 반영된 것이며, LFS의 최종 가중치를 의미하는 것은 아니다. 일반적으로 ER 지역의 소지역 통계 작성에 사용되는 보조정보는 주(Province) 단위의 노동력 통계 작성에 이용될 수 있는 보조정보에 비해 제한적이라 할 수 있다.

Drew et al.(1982)는 LFS에 알맞은 소지역 추정량을 찾기 위해 여러 종류의 추정량들을 비교하였다. 여기에서 사용된 소지역은 CD(Census Division) 지역이고 각 소지역에 대해 다음과 같은 세 개의 범주에 대한 보조정보를 추정과정에 이용하였다.

- (i) 연령 15~16세, 65세 이상 전체인구
- (ii) 연령 17~64세 여성인구
- (iii) 나이 17~64세 남성인구 .

소지역 a 를 포함하는 주(Province) 단위에서 $\hat{\beta}$ 을 구하여 합성추정량 $\hat{Y}_{SYN, GREG, a} = \hat{\beta} X_a$ 를 계산하였다. 여기에서 $\hat{\beta} = \sum_{i \in S} w_i y_i x_i' \left(\sum_{i \in S} w_i x_i x_i' \right)^{-1}$

이며, 주 단위에서 사용된 보조정보는 30개의 연령 및 성별그룹의 총 수, 각 ER 지역과 CMA 지역에 대한 인구 총계 등이다.

LFS에서 사용되고 있는 표본수 의존 추정량은 앞서 언급한 두 추정량의 가중평균

$$\hat{Y}_{SSD, a} = \lambda_a \hat{Y}_{GREG, a} + (1 - \lambda_a) \hat{Y}_{SYN, GREG, a}$$

을 취한다. 여기에서 가중치 λ_a 는

$$\lambda_a = \begin{cases} 1 & , \hat{N}_{e, a} \geq \delta N_a \\ \hat{N}_{e, a} / \delta N_a & , otherwise \end{cases}$$

이고, δ 값은 2/3를 사용하고 있다. 만약 소지역에 대한 직접추정량이 목표 요구정도를 만족하고 있다면 합성추정량 부분에 대한 가중치는 0이 된다.

현재 LFS에서 소지역의 취업률과 실업률에 대한 추정값은 표본수 의존 추정량으로 구해진 추정치의 3개월 간의 평균값이 이용되고 있다. 캐나다 LFS에서는 6개월의 표본순환 체계를 따르고 있기 때문에 3개월의 결과를 평균하게 되면 결과적으로 1/3표본 수를 늘리는 효과를 갖게 된다. 만약 조사시점 간에 표본들이 정확히 일치한다면 추정의 효율 측면에서 이러한 이득은 기대할 수 없다.

연속조사에서 추정량의 정확도를 향상시키기 위해 서로 다른 시점에서 조사된 몇 개의 조사 자료를 풀링(pooling)하는 경우를 흔히 볼 수 있다. 특히 시점이 다른 몇 차례의 조사결과를 결합하거나 이들의 평균을 구하는 것이 보통의 방법인데, 이러한 방법은 소지역 통계 작성 시 해당 소지역에 배당된 표본 크기가 매우 작아서 추정의 정확도가 크게 떨어지는 경우에 유용하다. 그러나 다른 시점의 조사결과들을 결합하여 산출되는 추정값은 개념상의 문제점을 항상 내재하고 있다.

대영역 내에서 추정된 소지역 추정값들의 합계는 대영역의 추정값과 일치해야 하지만, 대개의 경우 소지역 추정값들의 합계는 대영역의 추정값과 일치하지는 않는다. 따라서 최종적으로 총계를 일치시키는 다음과 같은 보정이 이루어진다. 소지역 a 에 대한 총계 Y_a 의 추정량을 \bar{Y}_a , 해당 소지역을 포함하는 대영역 A 에 대한 \bar{Y}_a 의 합계를 $\bar{Y}(A)$ 라 하자. 이때 소지역 추정값들은 다음 (3.56)식의 \bar{Y}_a^{ADJ} 를 통해 보정된다.

$$\bar{Y}_a^{ADJ} = \frac{\bar{Y}_a}{\bar{Y}(A)} \hat{Y}(A) \quad (3.56)$$

여기에서 $\bar{Y}(A) = \sum_{a \in A} \bar{Y}_a$ 이다.

(6) AK 복합추정

LFS에서 m 번째 달의 노동력 관련 특성치에 대한 총계를 Y_m 이라 하고, i 번째 패널에서 Y_m 에 대한 단순 비추정량을 $y_{m,i}$ 라 하자. 여기에서 i 는 m 번째 달에 조사되는 6개 패널 각각을 의미한다. $m-1$ 번째 달과 m 번째 달의 총계에 대한 변화량 $Y_m - Y_{m-1}$ 의 추정량 $d_{m,m-1}$ 은 6개의 패널 중 공통인 5개의 패널을 이용하여

$$d_{m,m-1} = \frac{\sum_{j=2}^6 (y_{m,j} - y_{m-1,j-1})}{5} \quad (3.57)$$

와 같이 정의된다고 하자.

이때 LFS 자료를 이용하여 추정되는 총계 Y_m 에 대한 AK복합추정량 y' 은 다음 (3.58)식의 형태로 주어질 수 있다.

$$y'_m = \frac{(1-K+A)y_{m,1}}{6} + \frac{(1-K-A/5)\sum_{j=2}^6 y_{m,j}}{6} + K(y'_{m-1} + d_{m,m-1}) \quad (3.58)$$

여기에서 K 와 A 는 상수들이고, K 는 $0 \leq K < 1$ 을 만족하는 값이다.

위의 (3.58)식은 AK 복합추정량의 일종이다. 한편, 위의 AK 복합추정량은 $A=0$ 인 경우에는 K 복합추정량으로 불리우며, $A=0, K=0$ 인 경우에는 6개의 패널 추정량들에 대한 평균인 $\bar{y}_m = \sum_{i=1}^6 y_{m,i}/6$ 의 값을 갖는다.

단순 비추정량 $y_{m,i}$ 에 대한 패널 간의 편향

$$\alpha_i = E(y_{m,i}) - Y_m \quad (3.59)$$

이 조사 달 m 과 서로 독립이라고 가정하면 AK 복합추정량 y' 의 편향은 다음과 같은 과정에 의해 유도될 수 있다.

(3.59)식으로부터 $E(y_{m,i}) = Y_m + \alpha_i$ 이고, $y_{m,i}$ 는 조사 달 m 과는 서로 독립이므로

$$\begin{aligned} E(\bar{y}_m) &= Y_m + \sum_{i=1}^6 \frac{\alpha_i}{6} \\ &= Y_m + \bar{\alpha} \end{aligned}$$

이 성립하며, LFS에서는 보통 $\bar{\alpha} = 0$ 이 가정된다. 한편, 식 (3.58)은 다음 (3.60)식과 같이 나타낼 수 있다.

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}) \quad (3.60)$$

여기에서

$$y_m = \frac{(1-K+A)y_{m,1}}{6} + \frac{(1-K-A/5)\sum_{j=2}^6 y_{m,j}}{6}$$

$$= (1-K)\bar{y}_m + \frac{A(y_{m,1} - \bar{y}_m)}{5}$$

따라서 y_m 의 기댓값은

$$E(y_m) = (1-K)(Y_m + \bar{\alpha}) + \frac{A(\alpha_1 - \bar{\alpha})}{5} \quad (3.61)$$

으로 주어지며, $\bar{\alpha} = 0$ 일 경우에는 다음 식으로 축약될 수 있다.

$$E(y_m) = (1-K)Y_m + \frac{A\alpha_1}{5}$$

한편, (3.57)식으로부터 $d_{m,m-1}$ 의 기댓값을 계산하면 다음 (3.62)식과 같다.

$$\begin{aligned} E(d_{m,m-1}) &= E\left(\frac{\sum_{j=2}^6 (y_{m,j} - y_{m-1,j-1})}{5}\right) \\ &= (Y_m - Y_{m-1}) + \frac{\alpha_6 - \alpha_1}{5} \end{aligned} \quad (3.62)$$

위와 같은 과정을 n 회의 이전 조사 달까지 반복 적용하면 AK 복합추정량 y' 은

$$\begin{aligned} y'_m &= y_m + Ky_{m-1} + K^2y_{m-2} + \cdots + K^{n-1}y_{m-n+1} + K^ny'_{m-n} \\ &\quad + Kd_{m,m-1} + K^2d_{m-1,m-2} + \cdots + K^nd_{m-n+1,m-n} \end{aligned} \quad (3.63)$$

와 같은 표현식으로 나타낼 수 있다. (3.63)식을 이용하여 y' 의 기댓값을 계산하면 다음 (3.64)식과 같이 주어진다.

$$\begin{aligned} E(y'_m) &= (1-K)(Y_m + KY_{m-1} + K^2Y_{m-2} + \cdots + K^{n-1}Y_{m-n+1}) \\ &\quad + K^n E(y'_{m-n}) + \left\{ (1-K)\bar{\alpha} + \frac{A(\alpha_1 - \bar{\alpha})}{5} \right\} \frac{(1-K^n)}{(1-K)} \\ &\quad + K(Y_m - Y_{m-1}) + K^2(Y_{m-1} - Y_{m-2}) + \cdots \end{aligned}$$

$$\begin{aligned}
& + K^n(Y_{m-n+1} - Y_{m-n}) + \left(\frac{\alpha_6 - \alpha_1}{5}\right) \left(\frac{K(1-K^n)}{1-K}\right) \\
& = Y_m + K^n\{E(y'_{m-n}) - Y_{m-n}\} \\
& \quad + \left\{ (1-K)\bar{\alpha} + \frac{A(\alpha_1 - \bar{\alpha})}{5} + \frac{K(\alpha_6 - \alpha_1)}{5} \right\} \left(\frac{1-K^n}{1-K}\right) \\
& = Y_m + K^n\{E(y'_{m-n}) - Y_{m-n}\} \\
& \quad + \left\{ (1-K - A/5)\bar{\alpha} + \frac{A\alpha_1}{5} + \frac{K(\alpha_6 - \alpha_1)}{5} \right\} \left(\frac{1-K^n}{1-K}\right)
\end{aligned} \tag{3.64}$$

여기에서 만약 n 이 충분히 크고, $\bar{\alpha} = 0$ 을 가정할 수 있다면, 위의 (3.64) 식은 다음과 같이 축약된다.

$$E(y'_m) = Y_m + \frac{A\alpha_1 + K(\alpha_6 - \alpha_1)}{5(1-K)} \tag{3.65}$$

또한 y'_m 의 편향은 월별 편향들과는 서로 독립이라고 가정하였기 때문에 이 전 조사달들과의 차이에 대한 변화량 $y'_m - y'_{m-r}$ 은 불편성을 만족한다. 즉,

$$E(y'_m - y'_{m-r}) = Y_m - Y_{m-r} \tag{3.66}$$

의 관계가 성립한다.

LFS의 연동표본교체 체계를 도시하면 <표3.22>와 같다.

<표3.22> LFS의 표본교체 체계

m	$m-1$	$m-2$	$m-3$	$m-4$	$m-5$	$m-6$	$m-7$	$m-8$	$m-9$	$m-10$	$m-11$
1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))	((2))
2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))
3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))
4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))
5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))
6	5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)

<표3.22>의 m 조사달과 $m-5$ 조사달 사이에서는 패널 간의 차 $i-j$ 의 값이 1과 5사이에 있을 경우, m 조사달의 i 패널($i=1,2,\dots,6$)은 $m-j$ 조사달의 $i-j$ 패널과 같다. m 조사달을 기준으로 바로 이전의 첫 번째 표본순환이 이루어지고 있는 6개월 간($m-1$ 에서 $m-6$)을 살펴보면, $6+i-j$ 의 값이 1과 6사이에 있을 경우, m 조사달의 i 패널은 $m-j$ 조사달의 $6+i-j$ 패널과 같다. 두 번째 표본순환이 이루어지고 있는 $m-7$ 조사달과 $m-12$ 조사달 사이에서는 $12+i-j$ 의 값이 1과 6사이에 있는 경우, m 조사달의 i 패널은 $m-j$ 조사달의 $12+i-j$ 패널과 같으며, 일반적으로 이전의 r 번째 표본순환에서 m 조사달의 i 패널은 $m-j$ 조사달의 $6r+i-j$ 패널과 같다.

따라서 AK 복합추정량 y' 의 분산 $V(y'_m)$ 은 위의 연동교체표본 체계에서 추정된 패널 추정량들의 분산공분산 값들을 포함하게 되며 $V(y'_m)$ 을 계산하기 위해 가정된 패널 추정량들의 분산공분산 구조는 다음과 같다.

$$(i) V(y_{m,i}) = \sigma^2, \quad i = 1, 2, \dots, 6,$$

$$(ii) Cov(y_{m,i}, y_{m-j,6r+i-j}) = \gamma_j^{(r)} \sigma^2, \quad i = 1, 2, \dots, 6,$$

$$j > 0, r \geq 0, 6 \geq 6r+i-j \geq 1,$$

$$\text{단, } \gamma_j^{(r)} = \begin{cases} \rho_j, & r=0 \text{ (i.e., } 6 > i-j \geq 1) \\ \gamma_j, & r=1 \text{ (i.e., } 6 \geq 6+i-j \geq 1) \\ 0, & r \geq 2 \text{ (i.e., } 6 \geq 6r+i-j \geq 1) \end{cases}$$

(iii) ρ_j 와 γ_j 는 정상성(stationary)을 만족한다. 즉 ρ_j 와 γ_j 는 j 의 함수이고, m 조사달과는 독립이다.

(iv) (ii)와 (iii)의 가정에 포함되지 않은 다음과 같은 공분산들은 0으로 가정한다.

$$Cov(y_{m,i}, y_{m,j}) = 0, \quad i \neq j,$$

$$Cov(y_{m,i}, y_{m-1,j}) = 0, \quad i = 1, j \neq 6,$$

$$Cov(y_{m,i}, y_{m-1,j}) = 0, \quad i \neq 1, j \neq i-1,$$

$$Cov(y_{m,i}, y_{m-g,j}) = 0, \quad g \geq 12.$$

추가적으로 AK 복합추정량 y' 의 월 추정값들이 매우 안정적이라면 $V(y'_{m-1}) = V(y'_m)$ 을 가정할 수 있다. 위의 (i)~(iv)의 가정과 y' 의 월 추정값들이 매우 안정적이라는 가정 하에서 복합추정량 y' 의 분산을 유도하면 다음과 같다.

먼저 (3.60)식의 AK 복합추정량 y' 에 대한 분산은 다음 (3.67)식과 같이 주어진다.

$$\begin{aligned} V(y'_m) = & [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1}) \\ & + 2KCov(y_m, y'_{m-1}) + 2K^2Cov(d_{m,m-1}, y'_{m-1})] / (1 - K^2) \end{aligned} \quad (3.67)$$

(3.63)식에서 m 을 $m-1$ 로, n 을 12로 대체하면 y'_{m-1} 은 다음과 같이 나타낼 수 있다.

$$y'_{m-1} = \sum_{g=1}^{12} (K^{g-1}y_{m-g} + K^g d_{m-g,m-g-1}) + K^{12}y'_{m-13} \quad (3.68)$$

(3.63)식을 (3.68)식에 대입하고 가정에 따라 0의 값을 갖는 항들을 제거하면 $V(y'_m)$ 은

$$\begin{aligned} V(y'_m) = & [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1}) \\ & + 2\sum_{g=1}^{12} K^g \{Cov(y_m, y_{m-g}) + KCov(d_{m,m-1}, y_{m-g}) \\ & + KCov(y_m, d_{m-g,m-g-1}) \\ & + K^2Cov(d_{m,m-1}, d_{m-g,m-g-1})\}] / (1 - K^2) \end{aligned} \quad (3.69)$$

와 같이 주어진다. 여기에서 $V(y_m)$, $V(d_{m,m-1})$, $Cov(y_m, d_{m,m-1})$ 은 (3.57)식과 (3.60)식으로부터 유도될 수 있고 계산결과는 다음과 같다.

$$V(y_m) = \left\{ \frac{(1-K)^2}{6} + \frac{A^2}{30} \right\} \sigma^2, \quad (3.70)$$

$$V(d_{m,m-1}) = \frac{2\sigma^2(1-\rho_1)}{5}, \quad (3.71)$$

$$\text{Cov}(y_m, d_{m,m-1}) = \frac{(1-K)(1-\rho_1)\sigma^2}{6} - \frac{A(1-\rho_1)\sigma^2}{30}, \quad (3.72)$$

$$\begin{aligned} \text{Cov}(y_m, y_{m-g}) &= \sigma^2 \rho_g I(g, 5) \{ (1-K)^2(6-g)/36 \\ &\quad + A(1-K)(g-3)/90 - gA^2/900 \} \\ &\quad + \sigma^2 \gamma_g \{ (1-K)^2(6-|g-6|)/36 \\ &\quad + A(1-K)(|g-6|-3)/90 - |g-6|A^2/900 \} \\ &\quad + \sigma^2 \gamma_g I(g, 6) I(6, g) A(1-K+A)/30, \end{aligned} \quad (3.73)$$

$$\begin{aligned} \text{Cov}(d_{m,m-1}, y_{m-g}) &= \sigma^2 (\rho_g - \rho_{g-1}) I(g, 5) \{ (1-K)(6-g)/30 \\ &\quad + gA/150 \} + \sigma^2 (\gamma_g - \gamma_{g-1}) \{ (1-K)(6-|g-6|)/30 \\ &\quad + |g-6|A/150 - (1-K+A)I(g, 6)/30 \}, \end{aligned} \quad (3.74)$$

$$\begin{aligned} \text{Cov}(y_m, d_{m-g, m-g-1}) &= \sigma^2 (\rho_g - \rho_{g+1}) I(g, 5) (1-K - \frac{A}{5}) \\ &\quad \times \left(\frac{5-g}{30} \right) + \sigma^2 (\gamma_g - \gamma_{g+1}) \{ (1-K - \frac{A}{5})(6 - I(6, g) \\ &\quad - |g-6|)/30 + AI(g, 5)/25 \}, \end{aligned} \quad (3.75)$$

$$\begin{aligned} \text{Cov}(d_{m,m-1}, d_{m-g, m-g-1}) &= \sigma^2 \{ (5-g)(2\rho_g - \rho_{g-1} - \rho_{g+1}) \\ &\quad \times I(g, 5) + (5-|g-6|)(2\gamma_g - \gamma_{g-1} - \gamma_{g+1}) \} / 25. \end{aligned} \quad (3.76)$$

단, $I(a, b) = \begin{cases} 1 & , \text{if } a \leq b \\ 0 & , \text{otherwise} \end{cases}$.

한편, $y'_m - y'_{m-1}$ 의 분산도 (3.60)식과 위의 결과들로부터 유도될 수 있다. 먼저 (3.60)식을

$$y'_m - Ky'_{m-1} = y_m + Kd_{m,m-1}$$

와 같이 나타내고 양변에 대해서 분산을 취하면 다음과 같은 결과를 얻을 수 있다.

$$(1 + K^2)V(y'_m) - 2KCov(y'_m, y'_{m-1}) = V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1})$$

$K \neq 0$ 인 경우를 고려해 보자. $Cov(y'_m, y'_{m-1})$ 는 위의 식으로부터 계산이 가능하다. 앞서 언급된 (3.70), (3.71) 및 (3.72)의 결과와

$$V(y'_m - y'_{m-1}) = 2V(y'_m) - 2Cov(y'_m, y'_{m-1})$$

의 관계를 이용한다면 $y'_m - y'_{m-1}$ 의 분산은 다음 (3.77)식과 같이 주어진다.

$$V(y'_m - y'_{m-1}) = \sigma^2 \{ A^2/30 - (1 - \rho_1)KA/15 + (1 - K)^2/6 + (1 - \rho_1)K(K + 5)/15 \} / K - (1 - K)^2 V(y'_m) / K \quad (3.77)$$

$K = 0$ 인 경우에 대해서는

$$y'_m = \left(1 - \frac{A}{5}\right) \bar{y}_m + \frac{Ay_{m,1}}{5},$$

$$Cov(y'_m, y'_{m-1}) = Cov\left\{\left(1 - \frac{A}{5}\right) \bar{y}_m + \frac{Ay_{m,1}}{5}, \left(1 - \frac{A}{5}\right) \bar{y}_{m-1} + \frac{Ay_{m-1,1}}{5}\right\} \quad (3.78)$$

을 이용하여 $y'_m - y'_{m-1}$ 의 분산 공식을 유도할 수 있고 결과는 다음 (3.79)식과 같이 주어진다.

$$V(y'_m - y'_{m-1}) = \sigma^2 \{ (1/15 + \rho_1/450 + \gamma_1/90)A^2 + 2(\rho_1 - \gamma_1)A/45 + 1/3 - (5\rho_1 + \gamma_1)/18 \} \quad (3.79)$$

위의 결과들에 대한 적용사례로써 캐나다 Ontario 지역의 LFS자료 (1980~1981년)에 근거한 추정결과를 소개하기로 한다. $V(y'_m)$ 의 표현식

에서 σ^2 , ρ_j 와 γ_j 는 그들의 추정값으로 대체하여 계산되었다. 캐나다 LFS에서 $j \geq 6$ 의 범위에 대해서는 접치는 패널들이 없기 때문에 이 범위에서는 ρ_j 값들이 존재하지 않으며, γ_j 도 $j \geq 12$ 의 범위에서는 값들이 존재하지 않는다. LFS자료에 근거하여 추정된 $\hat{\rho}_j$ 의 값이 다음 <표3.23>에 주어졌다. 여기에서 $\hat{\rho}_5$ 의 값은 표본으로부터 직접 계산할 수 없기 때문에 외삽법으로 계산된 값이다.

<표3.23> 특성치들에 대한 추정상관계수 $\hat{\rho}$ (1980-1981, Ontario)

구분	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$
노동력인구	0.843	0.782	0.717	0.674	0.631
취업자	0.852	0.779	0.709	0.664	0.619
농업인구	0.955	0.926	0.901	0.861	0.821
비농취업인구	0.861	0.791	0.724	0.678	0.632
실업자	0.580	0.445	0.334	0.286	0.238

예상했던 바와 같이 노동력인구, 취업자, 농업인구, 비농취업인구 및 실업자의 5가지의 관심영역에 대한 추정값 $\hat{\rho}_j$ 은 j 값이 증가함에 따라 감소하는 경향을 보인다. 특히 실업자를 제외한 모든 특성치들에 대해서 $\hat{\rho}_j$ 의 값들은 매우 큰 값을 나타낸다.

5가지의 노동력 관심영역에 대한 추정값 $\hat{\gamma}_j$ 는 다음 <표3.24>에 주어졌다. 여기에서 $\hat{\gamma}_5$ 는 내삽법에 의해 계산된 값이고, $\hat{\gamma}_{11}$ 은 외삽법에 의해 계산된 결과이다.

<표3.24> 특성치들에 대한 추정상관계수 $\hat{\gamma}$ (1980-1981, Ontario)

구분	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
노동력인구	0.16	0.14	0.13	0.13	0.14	0.14	0.13	0.13	0.12	0.12	0.13
취업자	0.16	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.14	0.15
농업인구	0.48	0.48	0.47	0.49	0.48	0.47	0.46	0.43	0.39	0.32	0.25
비농취업인구	0.18	0.15	0.15	0.16	0.16	0.17	0.17	0.17	0.17	0.16	0.17
실업자	0.14	0.07	0.08	0.06	0.06	0.05	0.05	0.06	0.08	0.14	0.07

일반적으로 각 특성치들에 대해서 j 가 증가할 때 γ_j 는 감소할 것으로 예측되지만 <표3.24>에서 γ_j 의 추정값들을 살펴보면 반드시 그러한 경향을 보이지는 않으며, 오히려 $\hat{\gamma}_{j+1} - \hat{\gamma}_j$ 의 값이 양의 값을 갖는 경우도 발생하는 것을 확인할 수 있다. 여기에서 $\hat{\gamma}_{j+1} - \hat{\gamma}_j$ 의 값이 양의 값을 갖는 경우에 대해서는 실제 상황으로 받아들이기 보다는 표본추출 오차에 기인하는 것으로 판단하는 것이 옳을 것 같다.

단순 비추정량의 추정분산에 대한 AK 복합추정량 및 K 복합추정량의 추정분산의 비를 계산한 상대효율 값들과 AK 복합추정량 또는 K 복합추정량의 분산을 최소화하는 최적상수 A 와 K 의 값들이 다음 <표3.25>와 <표3.26>에 주어졌다. 여기에서 분산에 대한 상대효율값들은 다음과 같이 정의되었다.

$$\text{상대효율} = \frac{\text{Var}(\text{단순비추정량})}{\text{Var}(AK\text{복합추정량 또는 } K\text{복합추정량})} \times 100$$

<표3.25>와 <표3.26>에서 최적 상수값들은 각 노동력 관심영역에 대해 분산을 최소화하는 값들로써 적절히 선택된 값들이다. <표3.25>의 결과를 살펴보면, 실업자를 제외한 나머지 노동력 관심영역들에 대해서 K 복합추정량은 단순 비추정량에 비해 약 18 ~ 21%의 효율이득이 발생하였고, AK 복합추정량은 약 26 ~ 30%의 효율이득이 발생하였다. 상대적으로 AK 복합추정량의 효율이득이 K 복합추정량의 효율이득보다 다소

높게 나타난다.

<표3.25> 추정량들의 상대효율 및 최적 상수($\gamma_i \neq 0$)

구 분	K 복합추정		AK 복합추정		
	최적K값	상대효율	최적K값	최적A값	상대효율
노동력인구	0.7	118.8	0.8	0.48	128.4
취업자	0.7	118.5	0.8	0.49	128.1
농업인구	0.8	120.6	0.8	0.38	126.9
비농취업인구	0.7	119.4	0.8	0.47	129.3
실업자	0.3	102.8	0.5	0.38	105.2

<표3.26> 추정량들의 상대효율 및 최적 상수($\gamma_i = 0$)

구 분	K 복합추정		AK 복합추정		
	최적K값	상대효율	최적K값	최적A값	상대효율
노동력인구	0.7	125.5	0.8	0.50	138.4
취업자	0.7	125.3	0.8	0.51	137.9
농업인구	0.8	167.3	0.9	0.46	187.9
비농취업인구	0.7	126.9	0.8	0.49	140.2
실업자	0.4	104.4	0.6	0.51	108.4

<표3.26>은 상대효율에 있어서 γ_j 의 효과를 알아보기 위해 γ_j 의 값들을 0으로 놓고 산출한 결과이다. <표3.25>와 <표3.26>을 비교해 보면 γ 의 값이 0이 아닌 양의 값을 취할 때 5개 노동력 관심영역들의 분산 값들이 $\gamma=0$ 인 경우보다 훨씬 큰 값을 가지며 효율이득도 그만큼 감소한다. 따라서 $\gamma_j > 0$ 인 실제상황에서 $\gamma_j = 0$ 이 가정된다면 효율이득이 과대 추정될 수 있다는 사실을 이들의 결과로부터 주목해야 한다.

앞서 언급한 바와 같이 복합추정량의 결점은 추정량의 편향 가능성에 있다. 따라서 편향된 추정량의 분산만을 비교하여 상대효율을 거론하는 것보다는 MSE 값들을 산출하여 효율이득을 비교하는 것이 보다 바람직

하다. AK 복합추정량 y' 의 편향은 (3.65)식에서와 같이 패널 간의 편향 α_i 를 포함한다. 추정편향 $\hat{\alpha}_i = y_{m,i} - \hat{Y}_m$ 이 α_i 의 불편추정량이고(즉, \hat{Y}_m 이 Y_m 의 불편추정량), 단순 비추정량 \bar{y} 이 Y_m 의 불편추정량이라고 가정하여 계산된 $\hat{\alpha}_i = y_{m,i} - \sum_{i=1}^6 y_{m,i}/6$ 값들이 다음 <표3.27>에 주어졌다.

노동력인구, 취업자, 비농취업인구의 3가지 관심영역에 대해서 $\hat{\alpha}_1$ 의 값은 다른 값들에 비해 상당히 큰 음의 값을 가지며, 다른 $\hat{\alpha}_i$ 의 값들은 대부분 양의 값을 갖는다.

<표3.27> 편향 α_i 의 추정결과(단위:천명)

구 분	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$
노동력인구	-135.3	39.8	41.1	31.1	15.4	7.9
취업자	-141.7	35.5	34.9	31.3	25.4	14.8
농업인구	-4.2	-2.6	2.2	-0.1	4.2	0.5
비농취업인구	-137.5	38.0	32.7	31.3	21.2	14.3
실업자	6.4	4.3	6.2	-0.1	-9.9	-6.9

5개의 노동력 관심영역들 각각에 대해서 분산 및 MSE 값을 최소화하는 최적상수 K , A 값들 및 이들 최적상수값들에 의해 계산된 분산과 MSE값들이 다음 <표3.28>에 주어졌다. MSE에 대한 상대효율값들은 다음 식을 이용하여 계산되었다.

$$\text{상대효율} = \frac{MSE(\text{단순비추정량})}{MSE(\text{AK복합추정량 또는 } K \text{ 복합추정량})} \times 100$$

<표3.28> 단순 비추정량, K 복합추정량, AK 복합추정량의 분산과 MSE 값들의 비교

(i) 노동력인구

노동력인구	단순 비추정	K 복합추정		AK 복합추정		(공통상수) $K = 0.4$ $A = 0.4$
		(min MSE) $K = 0$	(min Var) $K = 0.7$	(min MSE) $K = 0.7$ $A = 0.7$	(min Var) $K = 0.8$ $A = 0.5$	
월추정값(10^3)	4480.7	4480.7	4547.5	4484.4	4527.6	4481.7
분산(10^6)	432.0	432.0	363.8	358.1	336.5	391.8
편향(10^3)	0	0	66.8	3.7	46.9	1.0
$MSE(10^6)$	432.0	432.0	4284.5	371.5	2532.4	392.9
상대효율		100.0	9.0	116.3	17.1	110.0

(ii) 취업자

취업자	단순 비추정	K 복합추정		AK 복합추정		(공통상수) $K = 0.4$ $A = 0.4$
		(min MSE) $K = 0$	(min Var) $K = 0.7$	(min MSE) $K = 0.8$ $A = 0.9$	(min Var) $K = 0.8$ $A = 0.5$	
월추정값(10^3)	4186.0	4186.0	4259.0	4183.6	4240.3	4188.0
분산(10^6)	473.3	473.3	399.6	397.7	369.5	428.9
편향(10^3)	0	0	73.0	-2.4	54.3	2.0
$MSE(10^6)$	473.3	473.3	5732.2	403.2	3320.9	432.8
상대효율		100.0	8.3	117.4	14.3	109.4

(iii) 농업인구

농업인구	단순 비추정	K 복합추정		AK 복합추정		(공통상수) $K = 0.4$ $A = 0.4$
		(min MSE) $K = 0.6$	(min Var) $K = 0.8$	(min MSE) $K = 0.8$ $A = 0.6$	(min Var) $K = 0.8$ $A = 0.4$	
월추정값(10^3)	142.0	143.4	145.7	143.2	144.1	142.1
분산(10^6)	85.7	75.6	71.1	68.7	67.6	80.8
편향(10^3)	0	1.4	3.7	1.2	2.1	0.1
$MSE(10^6)$	85.7	77.6	85.1	70.2	71.8	80.8
상대효율		110.5	110.7	122.2	119.4	106.1

(iv) 비농취업인구

비농취업인구	단순 비추정	K복합추정		AK복합추정		(공통상수) K = 0.4 A = 0.4
		(min MSE) K = 0	(min Var) K = 0.7	(min MSE) K = 0.8 A = 0.9	(min Var) K = 0.8 A = 0.5	
월추정값(10^3)	4043.9	4043.9	4114.7	4041.6	4096.6	4045.8
분산(10^6)	498.9	498.9	417.8	418.0	385.9	452.8
편향(10^3)	0	0	70.8	-2.3	52.7	1.9
MSE(10^6)	498.9	498.9	5436.1	423.3	3161.7	456.4
상대효율		100.0	9.2	117.9	15.8	109.3

(v) 실업자

실업자	단순 비추정	K복합추정		AK복합추정		(공통상수) K = 0.4 A = 0.4
		(min MSE) K = 0.2	(min Var) K = 0.3	(min MSE) K = 0.4 A = 0.4	(min Var) K = 0.5 A = 0.4	
월추정값(10^3)	294.8	294.1	293.7	293.9	293.2	293.9
분산(10^6)	117.5	114.9	414.3	112.5	111.7	112.5
편향(10^3)	0	-0.7	-1.1	-0.9	-1.6	-0.9
MSE(10^6)	117.5	115.4	115.7	113.3	114.4	113.3
상대효율		101.9	101.6	103.7	102.7	103.7

여기에서 최적상수 K 는 0.0에서부터 0.9까지 0.1의 간격으로 선택된 10개의 값들 중에서 MSE 값을 최소화시키는 상수 값이다. AK 복합추정량에서 상수(K, A)는 $K(=0.0 \sim 0.9, \text{간격}=0.1)$ 와 $A(=0.0 \sim 1.0, \text{간격}=0.1)$ 값들의 조합에 의해 최적가중치가 결정되었다. K 복합추정량과 AK 복합추정량의 편향은 (3.65)식을 이용하여 계산하였다(α_1 과 α_6 대신 <표 3.27>에서 추정된 $\hat{\alpha}_1$ 과 $\hat{\alpha}_6$ 을 대입하여 계산).

<표3.28>에서 MSE를 최소화하는 상수값들에 의해 계산된 추정결과들을 살펴보자. K 복합추정량은 “농업인구”와 “취업자”에서 단순 비추정량에 비해 약간의 효율이득이 발생하였다. 분산을 최소화하는 상수값들에

의해 계산된 추정결과에서 “노동력 인구”, “취업자”, “비농취업인구”에 대해서는 매우 낮은 10%미만의 상대효율값을 갖는 것으로 나타났다. 이는 편향의 영향에 기인하는 것으로 나타났다.

MSE를 최소화하는 상수값들에서 계산된 AK 복합추정값은 “실업자”를 제외한 나머지 4개 노동력 관심영역에서 단순 비추정값에 비해 16 ~ 22%의 상대효율이익이 발생하였다. 4개 노동력 관심영역에 대한 편향 값은 매우 작게 나타났다. 분산을 최소화하는 상수값들에 의해 계산된 추정결과를 살펴보자. “노동력인구”, “취업자”, “비농취업인구”에 대해서는 K 복합추정량과 마찬가지로 편향의 영향에 의해 매우 낮은 상대효율값을 나타낸다.

결론적으로 “실업자”를 제외한 나머지 4개의 노동력 관심영역에서 AK 복합추정량의 상대효율이 K 복합추정량의 상대효율보다 높게 나타났다(MSE를 최소화하는 상수값들에 의해 계산된 추정값들을 기준). 따라서 주어진 LFS 자료에서는 K 복합추정량보다는 AK 복합추정량의 효율이 높을 것으로 결론지을 수 있다.

한편, K 복합추정량 또는 AK 복합추정량의 월별변화량 $y'_m - y'_{m-1}$ 은 $Y_m - Y_{m-1}$ 의 불편추정량이다. <표3.29>에 이들 월별변화량에 대한 추정결과가 주어졌다.

<표3.29> 월별변화량에 대한 복합추정량의 상대효율

구 분	K 복합추정		AK 복합추정			공통상수 K, A = 0.4
	최적K값	상대효율	최적K값	최적A값	상대효율	상대효율
노동력인구	0.9	146.6	0.9	0.1	147.9	113.3
취업자	0.9	151.0	0.9	0.1	152.3	114.1
농업인구	0.9	234.7	0.9	0.0	234.7	112.3
비농취업인구	0.9	154.0	0.9	0.1	155.2	114.1
실업자	0.4	106.0	0.6	0.2	106.4	102.9

월변화량에 대해서는 K 복합추정량 및 AK 복합추정량의 상대효율(단순 비추정량에 대한 상대효율)이 “실업자”를 제외한 나머지 관심영역에서 매우 높게 나타나며, “노동력인구”, “취업자”, “비농취업인구”의 영역에서는 약 46 ~ 55%의 효율이득이 발생하였고, “농업인구”영역에서는 매우 높은 약 135%의 효율이득이 발생하였다.

(7) 노동력 추정값의 분산추정

(가) 배경

캐나다 노동력 조사(LFS)는 캐나다 통계청에서 실시하는 가장 큰 규모의 월 단위 가구 조사로써 주로 전국 단위, 주 단위 및 주 내의 소지역 단위에 대한 다양한 노동력 특성에 대한 추정값들을 생산하고 있다. 캐나다의 LFS는 6개의 순환 패널을 갖는 층화 다단계 연동교체 표본설계를 따른다. 매년 인구 센서스 후 LFS는 표본설계에서 부분적인 보완이 이루어져 왔으며, 특히, 1981년에는 표본추출, 데이터 수집 및 추정 방법론 등에서 포괄적인 보완이 이루어졌다. 이 시기에 주 내의 소지역에 대한 추정치의 신뢰도를 높이기 위한 사후층화 비추정 절차가 새로이 마련되었다.

여기에서는 분산추정의 방법론에 대한 연구결과를 요약하였다. 과거 LFS의 분산추정은 Keyfitz 절차를 일반화한 Woodruff의 계산법이 이용되었다(Woodruff 1971). 이 방법은 Keyfitz 방법으로 불리우기도 한다. LFS에서는 다음 소개되는 세가지 유형의 지역들이 표본설계에 반영된다. 주요도시들로 구성되어 있는 SR지역(self-representing area), 소규모 도시들과 시골을 포함하는 NSR지역(non-self-representing area)과 군사지역 등과 같은 특수지역들이 이러한 유형의 지역들이다. NSR지역과 특수지역들에 대한 분산추정은 비추정방법에 Keyfitz 방법을 혼합하여 적용하였다. 이 단계 랜덤그룹 표본설계가 반영된 SR지역들에 대해서는 Rao, Hartley and Cochran(1962)과 Rao(1975)의 분산 추정량을 이용하였고, 이 방법을 Keyfitz 방법을 이용한 분산추정량과 비교하였다. 한편, 추정값들

에 대한 분산 추정량들을 비 보정(ratio adjustment) 분산추정량들과 편향 및 안정성의 측면에서 비교하였다. 또한 반복 수의 증가에 따른 Keyfitz 분산추정량의 영향도 살펴보았다.

(나) SR 설계에 대한 분산추정

① SR 설계(SR Design)

LFS 표본설계에서 SR 지역들은 이단계 랜덤그룹 설계방식을 취하며, 일단계 추출단위(PSU)들은 확률비례추출로 추출되며 이단계 추출단위들은 계통추출이 이루어진다. 하나의 층에 N 개의 일차추출단위(PSU)가 있고, j 번째 PSU에 대한 크기 측도를 x_j ($j=1, 2, \dots, N$), 거주단위의 수를 M_j , 층에서의 추출률을 $1/W$, 층으로부터 추출된 PSU의 수를 n 이라 하자. N 개의 PSU는 n 개의 그룹으로 랜덤하게 분할되고 i 번째 랜덤 그룹은 N_i 개의 PSU들을 포함한다. 여기에서 $\sum_{i=1}^n N_i = N$ 이다. 우선 다음의 수식을 정의하자.

$$p_j = \frac{x_j}{\sum_{i=1}^N x_i}, \quad j=1, 2, \dots, N$$

$$\delta_{ij} = \begin{cases} 1 & , \text{if the } j^{\text{th}} \text{ PSU is in the } i^{\text{th}} \text{ group} \\ 0 & , \text{otherwise} \end{cases}$$

이때 $\pi_j = \sum_{i=1}^n \delta_{ij} p_j$ 는 i 번째 랜덤 그룹의 상대적인 크기를 나타낸다.

$a_{ij} = \delta_{ij} W p_j / \pi_{ij}$, $r_{ij} = a_{ij} - [a_{ij}]$ 이라 하고, $\{r_{ij}, j=1, 2, \dots, N\}$ 가 내림차순으로 정렬되었다고 할 때, 계통추출의 추출간격 W_{ij} 는 다음과 같이 정의할 수 있다.

$$W_{ij} = [a_{ij}] + 1, \quad j=1, 2, \dots, R$$

$$= [a_{ij}], \quad j=R+1, \dots, N$$

여기에서 $R = \sum_{j=1}^N r_{ij}, \sum_{j=1}^N W_{ij} = W, i=1,2,\dots,n..$

하나의 PSU는 n 개의 랜덤 그룹 각각으로부터 추출물 W_{ij} 에 비례하는 확률로 추출된다. i 번째 랜덤 그룹으로부터 추출된 j 번째 PSU는 $1/W_{ij}$ 의 비율로 부차추출된다. 이때 전체 추출물은 $1/W$ 이 된다. 각각의 랜덤 그룹은 1에서 6까지의 패널로 할당되고, 랜덤 그룹의 수 n 은 보통 6의 배수이고 각각의 패널은 같은 수의 랜덤 그룹을 갖는다. 각 랜덤 그룹으로부터 하나의 PSU가 추출되므로 i 번째 랜덤 그룹으로부터 추출된 PSU에서 부차추출물은 $1/W_i$ 이 된다. 랜덤 그룹 i 에서 선택된 거주단위의 수는 m_i 로 나타내기로 한다.

② 분산 추정량(Variance Estimator)

하나의 층에 대한 특성치 y 의 총계에 관심이 있다고 하자. j 번째 PSU에서 k 번째 거주단위에 대한 y 값을 y_{ik} ($k=1,2,\dots,M_j$)라 할 때, 총

계 $Y = \sum_{j=1}^N \sum_{k=1}^{M_j} y_{ik}$ 는 $\hat{Y} = W \sum_{i=1}^n y_i$ 로 추정될 수 있다. 여기에서 y_i 는 i 번째

그룹에서 선택된 PSU로부터 추출된 m_i 개의 거주단위에 대한 y 값들의 합을 나타낸다. \hat{Y} 의 분산추정량은 다음과 같은 방법으로 추정될 수 있다.

(i) Keyfitz의 분산 추정량(1957)

과거 표본설계에서 이용하였던 총계 추정량에 대한 분산 추정공식은 다음 (3.80)식과 같다.

$$\hat{V}_1(\hat{Y}) = W^2 \left(\sum_a y_a - \sum_e y_e \right)^2 \quad (3.80)$$

여기에서 \sum_a 는 홀수 패널들에 대한 합, \sum_e 는 짝수 패널들에 대한 합을 나타낸다. 위의 (3.80)식을 일반화시킨 일반화 Keyfitz 분산 추정공식은

다음 (3.81)식과 같다.

$$\hat{V}_2(\hat{Y}) = W^2 \frac{n}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.81)$$

여기에서 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 이며, $\hat{V}_2(\hat{Y})$ 가 효율성이나 안정성 측면에서 $\hat{V}_1(\hat{Y})$ 보다 선호될 수 있다.

(ii) Rao, Hartley and Cochran의 분산 추정량(1962)

Rao, Hartley and Cochran의 분산 추정공식은 i 번째 그룹으로부터 추출된 m_i 개의 거주단위의 수가 고정되어있고, 거주단위들은 단순임의 추출이 가정된 상태에서 유도된다. 분산 추정공식은 다음과 같이 주어진다.

$$\hat{V}_3(\hat{Y}) = A \sum_{i=1}^n \pi_i \left(\frac{M_i}{m_i} \frac{y_i}{p_i} - \hat{Y} \right)^2 + \sum_{i=1}^n \frac{\pi_i}{p_i} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_i^2 \quad (3.82)$$

여기에서

$$A = \frac{\sum_{i=1}^n N_i^2 - N}{N^2 - \sum_{i=1}^n N_i^2},$$

$$s_i^2 = \frac{1}{m_i - 1} \sum_{k=1}^{m_i} (y_{ik} - \bar{y}_i)^2$$

이고, M_i 는 i 번째 그룹에서 선택된 PSU에 속해있는 거주단위들의 수이고 M_i 개의 거주단위들 중 m_i 개의 거주단위들이 계통추출로 추출되나 분산추정값은 단순임의추출 하에서 계산된다. i 번째 그룹에서 선택된 PSU로부터 추출된 k 번째 거주단위에 대한 y 값이 y_{ik} 이며 \bar{y}_i 는 $\bar{y}_i = y_i/m_i$ 로 주어진다.

$\pi_i/p_i = W/W_i$ 이고, $M_i/m_i = W_i$ 이므로 (3.82)의 분산 추정공식은 다음

(3.83)식과 같이 주어질 수 있다.

$$\hat{V}_3(\hat{Y}) = A \sum_{i=1}^n \pi_i \left(W \frac{y_i}{\pi_i} - \hat{Y} \right)^2 + W \sum_{i=1}^n \left(1 - \frac{m_i}{M_i} \right) M_i s_i^2 \quad (3.83)$$

(iii) Rao의 분산 추정량(1975)

Rao의 분산 추정공식에서는 m_i 개의 거주단위들이 단순임의추출로 추출되나, 표본크기 m_i 를 확률변수로 취급하여 분산 추정공식을 유도하였다. Rao의 분산 추정공식은 다음 (3.84)식과 같이 주어진다.

$$\begin{aligned} \hat{V}_4(\hat{Y}) &= A \sum_{i=1}^n \pi_i \left(W \frac{y_i}{\pi_i} - \hat{Y} \right)^2 + \sum_{i=1}^n \left\{ \frac{\pi_i^2}{p_i^2} - A \left(\frac{\pi_i}{p_i^2} - \frac{\pi_i^2}{p_i^2} \right) \right\} \frac{M_i^2 s_i^2}{m_i} \\ &\quad - \sum_{i=1}^n \frac{\pi_i}{p_i} M_i s_i^2 \\ &= \hat{V}_3(\hat{Y}) + W^2 \sum_{i=1}^n m_i s_i^2 \left\{ 1 - \frac{W_i}{W} \right\} - A \left(\frac{1}{\pi_i} - 1 \right) \end{aligned} \quad (3.84)$$

Rao의 분산 추정공식은 이 단계 표본추출에서 확률표본크기가 가정되기 때문에 음의 값이 나올 가능성이 있음에 주의해야한다.

③ Monte Carlo Study

네 가지의 분산 추정량들의 편향과 상대적인 안정성을 검토하기 위한 몬테카를로 연구가 수행되었다. 이용된 자료는 1981년 센서스 자료 중 조사지역 내의 약 20%의 계통추출 표본이며 Halifax의 CMA(Census Metropolitan Area) 지역으로부터 19개의 층에 대해 검토가 이루어졌다. 추출률 $1/W$ 은 0.04로 주어진다. 19개 층에 대한 PSU의 수, 추출된 PSU의 수, 거주단위들의 수 및 기대 표본크기가 <표3.30>에 주어졌다.

<표3.30> 몬테카를로 연구를 위해 이용된 층

층	거주단위 수	PSU 수	추출된 PSU 수	기대 표본크기
1	737	49	6	29.5
2	490	33	4	19.6
3	745	45	6	29.8
4	720	34	6	28.8
5	621	37	6	24.8
6	630	38	6	25.2
7	503	31	4	20.1
8	340	23	4	13.6
9	472	33	4	18.9
10	468	33	4	18.7
11	367	28	4	14.7
12	390	23	4	15.6
13	626	36	6	25.0
14	650	39	6	26.0
15	350	22	4	14.0
16	736	46	6	29.4
17	573	35	6	22.9
18	773	48	6	30.9
19	866	64	8	34.6
합계	11,056	697	100	442.3

몬테카를로 방법을 이용하여 각 층에서 1,000 개의 표본이 독립적으로 생성되었다. t 번째 몬테카를로 표본생성으로부터 층 h 에 대한 총계 Y_h 의 추정값을 \hat{Y}_{ht} ($h = 1, 2, \dots, 19, t = 1, 2, \dots, 1000$), \hat{V}_{jht} ($j = 1, 2, 3, 4$)를 \hat{Y}_{ht} 의 네 개의 분산추정량이라 하고 다음을 정의하였다.

$$Y = \sum_{h=1}^{19} Y_h,$$

$$\hat{Y}_t = \sum_{h=1}^{19} \hat{Y}_{ht}, \quad t = 1, 2, \dots, 1000$$

$$\hat{V}_{jt} = \sum_{h=1}^{19} \hat{V}_{jht}, \quad j = 1, 2, 3, 4$$

여기에서 \hat{Y}_t 는 t 번째 몬테카를로 표본생성에서 얻어지는 총계 Y 의 추정값, $\hat{V}_j (j=1, 2, 3, 4)$ 는 분산 추정값을 나타낸다.

몬테카를로 기대값과 분산을 각각 E^* 와 V^* 로 표기할 때 T 개의 몬테카를로 표본생성에 대한 기대값과 분산은 각각 다음과 같이 주어진다.

$$E^*(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \hat{\theta}_t,$$

$$V^*(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \{\hat{\theta}_t - E^*(\hat{\theta})\}^2$$

위의 정의를 이용하여 \hat{Y} 의 몬테카를로 분산 $V^*(\hat{Y})$ 과 분산추정량 \hat{V}_j 의 몬테카를로 기대값 $E^*(\hat{V}_j)$ 와 몬테카를로 분산 $V^*(\hat{V}_j)$ 를 얻을 수 있다.

분산추정량 \hat{V}_j 의 편향과 백분위 편향은 각각 다음과 같이 정의될 수 있다.

$$B_j = E^*(\hat{V}_j) - V^*(\hat{Y}),$$

$$PB_j = 100 \frac{B_j}{V^*(\hat{Y})}, \quad j=1, 2, 3, 4$$

이때 \hat{V}_j 의 평균제곱오차 MSE 는 다음과 같이 주어진다.

$$MSE_j = V^*(\hat{V}_j) + B_j^2, \quad j=1, 2, 3, 4$$

Keyfitz 분산추정량 \hat{V}_1 에 대한 \hat{V}_j 의 상대효율은 다음과 같이 정의할 수 있다.

$$Rel. Eff(\hat{V}_j \text{ vs. } \hat{V}_1) = \sqrt{\frac{MSE_1}{MSE_j}}, \quad j=1, 2, 3, 4$$

분산추정량들에 대한 상대편향과 효율에 대한 결과는 <표3.31>과 <표3.32>에 주어졌다.

<표3.31> 총계 추정값에 대한 분산추정량들의 백분위 편향

구 분	백분위 편향(PB_j)			
	\hat{V}_1	\hat{V}_2	\hat{V}_3	\hat{V}_4
취업인구	23.4	24.5	-4.7	-6.3
실업인구	6.3	6.6	3.7	1.2
노동력인구	24.2	25.2	-5.1	-6.7

<표3.32> \hat{V}_1 에 대한 $\hat{V}_2, \hat{V}_3, \hat{V}_4$ 의 상대효율

구 분	백분위 편향(PB_j)		
	\hat{V}_2	\hat{V}_3	\hat{V}_4
취업인구	1.51	3.22	3.11
실업인구	1.52	1.71	1.76
노동력인구	1.49	3.24	3.12

편향에 대해서는 분산추정량 \hat{V}_1 과 \hat{V}_2 가 유사하며, \hat{V}_3 와 \hat{V}_4 도 비슷한 결과를 보인다. 또한 \hat{V}_1 과 \hat{V}_2 의 편향은 취업인구와 노동력인구에서 비교적 큰 양의 편향값을 보이는 반면 \hat{V}_3 와 \hat{V}_4 의 편향은 상대적으로 작은 값을 나타낸다. 효율성 측면에서 \hat{V}_3 와 \hat{V}_4 가 서로 유사하며 \hat{V}_1 과 \hat{V}_2 에 비해서는 월등한 효율을 보인다. \hat{V}_1 과 \hat{V}_2 중에서는 \hat{V}_2 의 효율이 더 좋게 나타난다.

전체인구에 대한 비 추정값들의 분산추정량에 대한 결과가 다음 <표 3.33>과 <표3.34>에 주어졌다. 비 추정값들에 대한 분산추정량은 $\hat{V}_j^{(R)}$ ($j=1, 2, 3, 4$)로 표기하였다.

<표3.33> 총계 추정값에 대한 분산추정량들의 백분위 편향

구 분	백분위 편향(PB_j)			
	$\hat{V}_1^{(R)}$	$\hat{V}_2^{(R)}$	$\hat{V}_3^{(R)}$	$\hat{V}_4^{(R)}$
취업인구	3.7	4.3	-1.1	-3.1
실업인구	5.3	5.5	4.0	1.4
노동력인구	4.5	5.0	-0.5	-2.5

<표3.34> $\hat{V}_1^{(R)}$ 에 대한 $\hat{V}_2^{(R)}$, $\hat{V}_3^{(R)}$, $\hat{V}_4^{(R)}$ 의 상대효율

구 분	백분위 편향(PB_j)		
	$\hat{V}_2^{(R)}$	$\hat{V}_3^{(R)}$	$\hat{V}_4^{(R)}$
취업인구	2.13	2.59	2.52
실업인구	1.57	1.71	1.76
노동력인구	2.08	2.56	2.51

$\hat{V}_1^{(R)}$ 과 $\hat{V}_2^{(R)}$ 의 편향이 취업인구와 노동력인구에서 \hat{V}_1 과 \hat{V}_2 에 비해 훨씬 작아진 사실을 확인할 수 있다. $\hat{V}_3^{(R)}$ 과 $\hat{V}_4^{(R)}$ 의 편향도 취업인구와 노동력인구에서 \hat{V}_3 와 \hat{V}_4 보다 작은 값을 가지며 실업인구에서는 거의 변화가 발생하지 않았다.

몬테카를로 표본을 이용하여 네 가지 분산추정량들의 비 보정 (ratio-adjustment) 추정값들에 대한 95% 신뢰구간을 살펴보았다. 추정값들의 95% 신뢰구간에 대한 포함비율이 다음 <표3.35>에 주어졌다.

<표3.35> 비보정을 갖는 총계 추정값들의 95% 신뢰구간 포함비율

구 분	포함 비율			
	$\hat{V}_1^{(R)}$	$\hat{V}_2^{(R)}$	$\hat{V}_3^{(R)}$	$\hat{V}_4^{(R)}$
취업인구	93.6	95.4	94.6	94.2
실업인구	94.3	95.1	95.3	95.0
노동력인구	93.2	95.3	94.6	94.2

취업인구, 실업인구 및 노동력인구의 모든 부분에서 네 가지 분산추정량들의 수행결과가 모두 적합한 것으로 나타났다.

편향의 관점에서는 비 보정 추정값들의 분산추정량들이 서로 상이한 결과를 나타내지는 않는다. 상대효율에서는 $\hat{V}_3^{(R)}$ 과 $\hat{V}_4^{(R)}$ 이 $\hat{V}_2^{(R)}$ 에 비해 효율이 높은 것으로 나타났다. 한편, $\hat{V}_1^{(R)}$ 의 자유도는 19 이고(각 층별로 1개의 자유도를 가짐) $\hat{V}_3^{(R)}$ 은 각 PSU가 하나의 반복으로 처리되어 81 개의 자유도를 갖게 된다. 따라서 비 보정 추정값들에 대한 Keyfitz 분산추정량은 반복 수를 증가시키면 추정량의 안정성을 확보할 수 있다.

④ 반복 수를 갖는 Keyfitz 분산 추정량

Keyfitz 방법의 효율성을 높이기 위해 6 개의 순환 패널들이 반복표본들로 채택되었다. 6 개의 순환 패널을 반복표본으로 이용한 분산 추정값들과 과거 표본설계를 이용한 2 개의 반복표본으로부터 계산된 분산 추정값들이 비교되었다. 순환 패널이 반복표본으로 처리됨에 따라 기인된 중요한 관심 사항은 패널 편향으로부터 발생할 수 있는 분산 추정값들의 증가부분이다. 이러한 부분을 살펴보기 위해 '85년 3월부터 '87년 2월까지의 24개월의 LFS 자료가 이용되었다. 취업인구, 실업인구, 노동력인구의 24개월에 대한 분산 추정값들에 대한 평균과 표준편차를 계산하였다. 2 개의 반복표본과 6 개의 반복표본 하에서 얻어진 분산들에 대한 평균과 표준편차의 비는 24 개의 CMA(Census Metropolitan Area) 지역들의 평균으로 산출하였고 다음 <표3.36>에 주어졌다.

<표3.36> 단순임의 분산추정값의 비교(LFS의 CMA지역 자료)

구 분	분산의평균에 대한 비 평균(average ratio) : 2 vs. 6 반복	분산의 표준편차에 대한 비 평균(average ratio) : 2 vs. 6 반복
취업인구	0.997	1.813
실업인구	0.995	1.515
노동력인구	1.003	1.833

6개의 반복표본을 이용한 분산이 2개의 반복표본의 분산보다 작은 값을 나타낸다. 순환 패널을 반복으로 채택할 경우 분산 추정값들의 편향에 거의 영향을 미치지 않으며, 6개의 반복표본을 이용한 분산이 2개의 반복표본보다는 훨씬 안정적임을 확인할 수 있다. 즉 Keyfitz 방법에서 6개의 순환 패널을 반복으로 사용할 경우 심각한 편향은 발생하지 않으며 2개의 반복표본을 이용했을 경우보다는 효율이 증가됨을 확인할 수 있다.

(다) 비 추정값 탐색을 위한 분산추정방법

① LFS에서 비 추정방법

과거 LFS에서는 사후층화 비 추정방법이 이용되었다. 무응답을 보정하기 위한 일종의 설계 가중치인 부차 가중치가 LFS 목표 모집단의 추정치들에 대해 비 보정되었다. 이러한 비 추정방법은 주 지역의 특성치에 대한 추정의 신뢰도를 높이는 결과를 보였으나 주 내의 소지역들에 대해서는 문제점을 안고 있었다. 주 내의 ER(Economical Region) 지역과 CMA(Census Metropolitan Area) 지역들에 대한 추정의 정확도를 높이기 위해 탐색적인 비 추정 절차가 채택되었다.

탐색적 추정절차는 보정값들의 수열을 통해 수행된다. 먼저 부차가중치가 주 내의 소지역의 인구를 참조하여 보정되며, 이 후 성별-연령대별 범주를 반영한 주 수준의 보정값이 최종 가중치에 적용된다. 이러한 절차는 한번 더 수행되어 한 쌍의 가중치가 추가적으로 생성된다. W_0 를 부차가중치라 하고 (W_1, W_2) 와 (W_3, W_4) 를 2회 반복으로부터 생성된 가중치들의 쌍이라 하자. 노동력 특성값들은 W_4 를 이용하여 추정된다. 주 지역의 성별-연령대별 그룹들에서 주변 총계 W_4 는 상응하는 그룹들의 외부 인구 추정치와 정확히 일치하나 주 내의 소지역에 대해서는 반드시 그렇지 않다. 그러나 그 차이는 매우 작게 나타난다.

② 1회 반복 비 추정값에 대한 분산공식

1회 반복 비 추정값들에 대한 분산공식을 유도하면 다음과 같다. 여기에서 적용되는 기본적인 방법론은 선형적인 형태의 부차 가중치를 얻을 때까지 테일러 전개 근사식을 연속적으로 적용하는 방법이다. 세부적인 유도과정을 소개하면 다음과 같다.

$Y^{(0)}, Y^{(1)}, Y^{(2)}$ 를 주 지역에서 W_0, W_1, W_2 에 근거하여 추정된 노동력 특성값 y 의 추정값들이라 하자. 이때 $Y^{(2)}$ 는 다음 (3.85)식과 같이 주어질 수 있다.

$$Y^{(2)} = \sum_a \frac{Y_a^{(1)}}{P_a^{(1)}} P_a \quad (3.85)$$

여기에서 $Y_a^{(1)}$ 은 주 지역에서 성별-연령대별 그룹 a 에 대한 특성치 y 의 W_1 가중추정값, $P_a^{(1)}$ 은 주 지역에서 성별-연령대별 그룹 a 에 대한 인구의 W_1 가중 추정값, P_a 는 주 지역에서 성별-연령대별 그룹 a 에 대한 인구의 외부 추정치를 나타낸다.

F_a 를 $F_a = Y_a^{(1)}/P_a^{(1)}$ 이라 할 때, $(E(Y_a^{(1)}), E(P_a^{(1)}))$ 에서 F_a 에 대한 일계 테일러 근사식을 구하면 다음과 같이 주어진다.

$$F_a \doteq \frac{E(Y_a^{(1)})}{E(P_a^{(1)})} + \frac{1}{E(P_a^{(1)})} \{Y_a^{(1)} - E(Y_a^{(1)})\} - \frac{E(Y_a^{(1)})}{\{E(P_a^{(1)})\}^2} \{P_a^{(1)} - E(P_a^{(1)})\}$$

이때 $Y^{(2)}$ 의 분산에 대한 테일러 근사식은 다음 (3.86)식과 같이 주어질 수 있다.

$$V(Y^{(2)}) = V\left(\sum_a F_a P_a\right) \doteq V\left\{\sum_a \frac{P_a}{E(P_a^{(1)})} (Y_a^{(1)} - R_{Y_a^{(1)}} P_a^{(1)})\right\} \quad (3.86)$$

여기에서 $R_{Y_a}^{(1)} = E(Y_a^{(1)})/E(P_a^{(1)})$ 을 나타낸다.

다음으로 W_1 가중 추정값 $Y_a^{(1)}$ 과 $P_a^{(1)}$ 은 W_0 가중 추정값의 항으로 다음 (3.87)식과 같이 나타낼 수 있다.

$$\begin{aligned} Y_a^{(1)} &= \sum_s \frac{Y_{sa}^{(0)}}{P_s^{(0)}} P_s \\ P_a^{(1)} &= \sum_s \frac{P_{sa}^{(0)}}{P_s^{(0)}} P_s \end{aligned} \quad (3.87)$$

여기에서 s 는 CMA 또는 ER 지역을 나타내며, P_s 는 지역 s 의 인구를 나타낸다. (3.87)식의 $Y_a^{(1)}$ 과 $P_a^{(1)}$ 을 (3.86)식에 대입하여 W_0 가중 추정값의 비에 대한 일계 테일러 근사식을 구하면 다음 (3.88)식과 같이 주어질 수 있다.

$$\begin{aligned} V(Y^{(2)}) &\doteq V\left[\sum_a \frac{P_a}{E(P_a^{(1)})} \sum_s \frac{P_s}{E(P_s^{(0)})} \{ (Y_{sa}^{(0)} - R_{Y_a}^{(0)} P_s^{(0)}) \right. \\ &\quad \left. - R_{Y_a}^{(1)} (P_{sa}^{(0)} - R_{P_a}^{(0)} P_s^{(0)}) \} \right] \end{aligned} \quad (3.88)$$

여기에서

$$\begin{aligned} R_{Y_a}^{(0)} &= \frac{E(Y_{sa}^{(0)})}{E(P_s^{(0)})}, \\ R_{P_a}^{(0)} &= \frac{E(P_{sa}^{(0)})}{E(P_s^{(0)})} \end{aligned}$$

이다.

위의 (3.88)식은 다음 (3.89)식과 같이 축약된 형태로 다시 표현할 수 있다.

$$\begin{aligned} V(Y^{(2)}) &\doteq V\left\{ \sum_s \sum_{h \in s} \sum_{i=1}^{n_h} \sum_a (Z_{Y_{sh}i}^{(1)} - R_{Y_a}^{(1)} Z_{P_{sh}i}^{(1)}) \right\} \\ &= V\left(\sum_s \sum_{h \in s} \sum_{i=1}^{n_h} D_{sh}^{(0)} \right) \end{aligned} \quad (3.89)$$

여기에서

$$D_{shi}^{(0)} = \sum_a (Z_{Y_{sa}^{(1)}} - R_{Y_a^{(1)}} Z_{P_{sa}^{(0)}}),$$

$$Z_{Y_{sa}^{(0)}} = \frac{P_a}{E(P_a^{(1)})} \frac{P_s}{E(P_s^{(0)})} (Y_{shia}^{(0)} - R_{Y_a^{(0)}} P_{shi}^{(0)}),$$

$$Z_{P_{sa}^{(0)}} = \frac{P_a}{E(P_a^{(1)})} \frac{P_s}{E(P_s^{(0)})} (P_{shia}^{(1)} - R_{P_a^{(0)}} P_{shi}^{(0)})$$

이고, h 는 s 에 속하는 층을 나타내며, i 는 층 h 에서 반복을 나타낸다.

식 (3.89)에서 $\left\{ \sum_{i=1}^{n_h} D_{shi}^{(0)} \right\}$ 는 부차가중치들에 의해 결정되므로 독립성을 가정할 수 있다. 따라서 식 (3.89)은 다음 (3.90)식과 같이 표현될 수 있다.

$$V(Y^{(2)}) \doteq V \left(\sum_{h \in s} \sum_{i=1}^{n_h} D_{shi}^{(0)} \right) \quad (3.90)$$

$$= V \left(\sum_h \sum_{i=1}^{n_h} \sum_{s \in h} D_{shi}^{(0)} \right)$$

여기에서 $\sum_{s \in h}$ 는 층 h 를 포함하고 있는 주 내의 모든 소지역들에 대한 합

을 나타낸다. $D_{hi}^{(0)}$ 를 $D_{hi}^{(0)} = \sum_{s \in h} D_{shi}^{(0)}$ 와 같이 정의하면, 위의 (3.90)식은

다음과 같이 주어진다.

$$V(Y^{(2)}) \doteq V \left(\sum_h \sum_{i=1}^{n_h} D_{hi}^{(0)} \right) \quad (3.91)$$

여기에서 $\left\{ \sum_i D_{hi}^{(0)} \right\}$ 는 부차가중치에 의해 결정되므로 이 변수들은 독립성

을 가정할 수 있으며, 분산은 다음 식으로부터 추정될 수 있다.

$$\hat{V}(Y^{(2)}) \doteq \sum_h \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (D_{hi}^{(0)} - \bar{D}_h^{(0)})^2 \quad (3.92)$$

여기에서 $\bar{D}_h^{(0)} = \frac{1}{n_h} \sum_{i=1}^{n_h} D_{hi}^{(0)}$ 이다. 그러나 이 표현식에서는 기대값이 포함되어있고 이러한 값들은 미지의 값이므로 추정값으로 대체하여 분산의 추정값을 근사적으로 계산할 수 있으며 이를 이용한 분산 추정값은 최종적으로 다음 (3.93)식과 같이 주어진다.

$$\hat{V} \doteq \sum_h \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (D_{hi}^{(2)} - \bar{D}_h^{(2)})^2 \quad (3.93)$$

여기에서

$$D_{hi}^{(2)} = \sum_{s \in h} D_{shi}^{(2)}$$

$$\bar{D}_h^{(2)} = \frac{1}{n_h} \sum_{i=1}^{n_h} D_{hi}^{(2)}$$

$$D_{shi}^{(2)} = \sum_a (Z_{Y_{sa}}^{(2)} - R_{Y_s}^{(2)} Z_{P_{sa}}^{(2)}),$$

$$\begin{aligned} Z_{Y_{sa}}^{(2)} &= \frac{P_a}{P_a^{(1)}} \frac{P_s}{P_s^{(0)}} \left(Y_{shia}^{(0)} - \frac{Y_{sa}^{(0)}}{P_s^{(0)}} P_{shi}^{(0)} \right) \\ &= Y_{shia}^{(2)} - \frac{P_{shi}^{(0)}}{P_s^{(0)}} Y_{sa}^{(2)} \end{aligned}$$

$$\begin{aligned} Z_{P_{sa}}^{(2)} &= \frac{P_a}{P_a^{(1)}} \frac{P_s}{P_s^{(0)}} \left(P_{shia}^{(0)} - \frac{P_{sa}^{(0)}}{P_s^{(0)}} P_{shi}^{(0)} \right) \\ &= P_{shia}^{(2)} - \frac{P_{shi}^{(0)}}{P_s^{(0)}} P_{sa}^{(2)} \end{aligned}$$

$$R_{Y_s}^{(2)} = \frac{Y_a^{(1)}}{P_a^{(1)}} = \frac{P_a}{P_a^{(1)}} \frac{Y_a^{(1)}}{P_a} = \frac{Y_a^{(2)}}{P_a}$$

위의 분산 추정공식 (3.93)식은 노동력 특성값들의 W_2 가중 추정값들에 대한 추정공식이며 W_0 와 W_2 의 두 가중치의 값을 요구한다.

③ 2회 반복 비 추정값의 분산 추정

2회 반복 비 추정값에 대한 분산공식은 앞에서 언급한 내용을 응용하여 테일러 급수전개의 연속적인 적용으로 얻을 수 있다. 그러나 2회 반복에 기인한 분산 추정공식은 매우 복잡한 형태를 취하기 때문에 1회 반복에 기인한 분산 추정공식이 오히려 합리적일 수 있다. 1회 반복 분산 공식은 한 쌍의 가중치 (W_0, W_2)를 이용한다. 여기에서는 (W_0, W_2) 대신에 (W_0, W_4)를 이용하였다. W_2 보다는 W_4 가 노동력 추정값들의 CV값들에 강한 영향력을 주지 않기 때문이다. 주 지역인 Nova Scotia 지역의 1981년 센서스 자료를 이용하여 몬테카를로 시뮬레이션이 수행되었다. 각각의 몬테카를로 표본에서 LFS 표본설계가 매 단계의 표본 추출을 통해 검증되었다. 1,000개의 몬테카를로 표본들이 독립적으로 추출되었다. 각각의 몬테카를로 표본들에 대해 2회 반복표본 비 추정값($Y^{(4)}$), 1회 반복 분산 추정량과 이의 CV 추정값을 이용한 분산 추정값($\hat{V}(Y^{(4)})$)과 95% 신뢰구간($Y^{(4)} \pm 1.96\sqrt{\hat{V}(Y^{(4)})}$)을 주 지역과 주 지역 내의 소지역들에 대해 계산하였다. 또한 1,000개의 CV값들의 평균을 계산하여 실제값과 매우 유사한 몬테카를로 CV값들과 비교하였다. 결과는 <표3.37>에 주어졌다. <표3.37>의 모든 셀에 대해서 CV값의 차이는 8% 미만으로 나타났다. 21개의 셀 중 13개의 셀에서는 4% 미만의 값을 갖는다. 한편 신뢰구간의 포함범위는 <표3.38>에 주어졌다. 취업인구와 노동력인구에 대한 95% 신뢰구간의 포함범위는 만족한 값을 보이나 실업인구에 대해서는 다소 낮은 포함범위를 나타내나 여전히 받아들일만한 결과를 보인다.

<표3.37> 1회 반복 분산추정량의 평균 CV값과 몬테카를로 CV값

구 분	ER 210	ER 220	ER 230	ER 240	ER 250	CMA Halifax	Province (Nova Scotia)
Average CV's							
취업인구	3.52	3.46	3.14	3.05	1.96	2.01	1.08
실업인구	10.36	12.28	13.13	13.43	10.35	10.55	5.27
노동력인구	2.98	3.17	2.85	2.73	1.77	1.83	0.91
Monte Carlo CV's							
취업인구	3.48	3.35	2.95	2.86	1.97	1.99	1.11
실업인구	10.90	12.71	13.28	13.37	11.12	11.31	5.59
노동력인구	2.76	3.08	2.76	2.53	1.72	1.74	0.92

<표3.38> 1회 반복 분산추정량에 의한 95% 신뢰구간의 포함범위

구 분	ER 210	ER 220	ER 230	ER 240	ER 250	CMA Halifax	Province (Nova Scotia)
취업인구	94.5	92.8	94.0	94.7	94.7	94.9	92.5
실업인구	92.1	90.7	91.4	91.8	92.7	92.7	93.1
노동력인구	96.2	93.0	93.6	95.2	95.2	96.0	94.0

(라) 요약

비 보정을 하지 않은 추정값들의 Keyfitz 분산 추정법은 매우 큰 양의 편향을 가지며 효율성도 크게 떨어진다. 반면에 탐색적 비 보정 추정방법은 상대적으로 작은 편향을 가지며 효율성도 크게 향상되는 것으로 확인되었다. 이 논문에서 소개된 비 보정 추정값들에 대한 분산 추정방법들은 무시할 수 있을 정도의 작은 편향을 갖는다. 한편 Keyfitz 방법은 반복수를 증가시킬 경우 다른 분산 추정방법에 비해 효율을 크게 향상시킬 수 있었다. LFS 자료에서 6개의 순환 패널을 반복으로 취급하여 Keyfitz 방법을 적용시켜본 결과 순환 패널 편향에 기인한 Keyfitz 추정분산의 편향은 발생하지 않았다. Keyfitz 방법에 의해 유도된 1회 반복 분산공식은 2회 반복의 탐색적 비 추정값들에 대해서 매우 합리적인 분산 추정값들을 제공하며 신뢰구간에 대한 포함범위도 좋은 특성을 나타낸다.

3.3 영국

3.3.1 서론

정부 공식통계는 경제 정책, 자원 분배 및 정책 결정 등에 참고 자료로 이용된다. 대 영역에 대해서는 공식 통계의 정보가 이용자들에게 제공되나 소지역에 대해서는 그렇지 못한 실정이다. 최근 영국 내에서는 소지역 통계 작성에 대한 요구가 꾸준히 제기되고 있고 특히 노동시장 동향에 대한 측도 개발이 시급히 요구되고 있다. 노동력 조사(LFS)는 노동시장 정보 파악에 중요한 역할을 담당하고 있으나 직접조사 추정값들은 소지역 추정에는 한계를 안고 있다. 이 논문은 LFS로부터 소지역 추정의 신뢰도를 향상시킬 수 있는 모형 기반 추정량들을 소개한다.

영국의 LFS는 세 달을 주기로 연속 조사가 실시된다. LFS는 약 60,000 조사가구 단위를 갖는 대규모 조사로써 16세 이상의 약 150,000 명의 인구에 대해 조사가 이루어진다. 영국의 LFS는 국제노동기구(ILO)의 요구조건을 만족하도록 표본설계되어 있으며 이를 통해 실업통계가 작성된다. LFS 표본설계에서 표본은 단순임의추출로 추출되며 주로 국가 수준의 추정값을 생산하도록 설계되어 있다. 일년에 한번 소지역 단위인 UA지역(Unitary Authority)과 LAD지역(Local Authority)에 대한 추정값들이 작성된다. 이 연구에서 소개되는 내용은 현재 영국 통계국(ONS)에서 진행하고 있는 소지역 추정법과 밀접한 관계가 있다.

소지역 추정법은 소지역에 대한 직접 조사 추정값들이 신뢰성에 문제가 있거나 계산될 수 없을 때 이용할 수 있는 통계적인 기법으로써 인근 지역의 보조정보를 빌려 소지역의 특성값을 추정하는 간접 추계 방법이다. 이 연구에서 이용한 주요 보조정보는 실업보험을 청구한 사람들의 수이다. 실업보험 자료는 행정 시스템에 의해 획득되기 때문에 표본오차가 없고 지역적 범주로 또는 성별-연령대별 범주들로 다양하게 분류될 수 있다. 실업보험 지급 청구자 수와 ILO 실업자 수와는 시기에 따라 약간의

차이는 있지만 강한 상관성을 나타낸다. 시기에 따라 발생하는 차이는 주로 행정 시스템의 변경 또는 경제 사이클의 변화 등에 기인한다. ONS는 Southampton 대학과 연계하여 LFS자료와 실업보험 청구자 수의 자료를 결합하여 소지역 추정값의 신뢰성을 확보할 수 있는 연구를 진행하고 있으며, 특히 UA 또는 LAD 지역에 대한 실업자 수를 추정하는 SPREE 방법(Purcell and Kish, 1980), 로지스틱 모형에 근거한 일종의 변형된 Fay-Herriot 방법(1979)과 Multi-level 모형화 방법(Goldstein, 1995)과 같은 세 가지 추정방법에 대해 연구를 진행하고 있다. SPREE 방법보다는 로지스틱 모형에 근거한 변형된 Fay-Herriot 방법과 Multi-level 모형화 방법이 더 좋은 효율을 나타내며, 여기에서는 변형된 Fay-Herriot 방법에 초점을 맞추어 소개한다.

3.3.2 소지역 추정 방법

앞으로 소개되는 수식에서 첨자 i 와 j 는 각각 UA 지역과 LAD 지역에 대한 성별-연령대별 그룹을 나타내며, 첨자 g 와 h 는 각각 UA 지역과 LAD 지역들을 나타낸다. 표본 자료는 LFS 추정값들과 각 지역 내에서 성별-연령대별 그룹으로 분류된 각 셀들에 대한 실업보험 청구자 수의 자료들로 이루어져 있다.

N_{ig} 를 셀 (i, g) 에서의 인구 총계, U_{ig} 를 같은 셀에서의 실업자 총계라 할 때, 이 셀에서의 실업률은 $Z_{ig} = U_{ig}/N_{ig}$ 이다. 일반적으로 실업률 Z_{ig} 는 g 번째 지역의 특성값들에 의해 결정된다. Z_{ig} 의 기대값과 분산을 각각

$$E(Z_{ig}) = \pi_{ig} ,$$

$$Var(Z_{ig}) = \frac{\pi_{ig}(1 - \pi_{ig})}{N_{ig}}$$

라 하자. g 번째 지역의 특성값들이 $E(Z_{ig})$ 의 값에 미치는 영향을 열거하기 위해 로지스틱 모형이 이용되었다. 이용된 로지스틱 모형은

$$\text{logit}(\pi_{ig}) = x_{ig}^t \beta$$

이다. 여기에서 벡터 x_{ig} 는 소지역 g 에서 i 번째 성별-연령대별 그룹에 대한 속성들을 나타내며 알고있는 값이다.

N_{ig}^* 를 셀 (i, g) 에서의 인구 총계에 대한 LFS 추정값이라 하고, U_{ig}^* 를 실업자 수에 대한 추정값이라 할 때, LFS 실업률 추정값은 $Z_{ig}^* = U_{ig}^*/N_{ig}^*$ 로 나타낼 수 있다. 성별-연령대별 그룹들이 합리적으로 정의되어 그룹 내에서 추출된 조사단위들에 대한 표본 가중치들에서 변동이 거의 발생하지 않는다고 가정할 수 있다면 셀 (i, g) 에서의 실업률에 대한 LFS 추정값들은 표본 실업률로 근사될 수 있다. LFS 표본은 단순임의추출 표본이므로 다음 식들이 성립한다.

$$E(Z_{ig}^* | Z_{ig}) = Z_{ig}, \quad (3.94)$$

$$\begin{aligned} V(Z_{ig}^* | Z_{ig}) &= \frac{N_{ig} - n_{ig}}{N_{ig} - 1} \frac{Z_{ig}(1 - Z_{ig})}{n_{ig}} \\ &= \frac{Z_{ig}(1 - Z_{ig})}{n_{ig}^*} \end{aligned} \quad (3.95)$$

여기에서

$$n_{ig}^* = \frac{n_{ig}^o (N_{ig}^* - 1)}{N_{ig}^* - n_{ig}^*}$$

이고, n_{ig} 는 셀 (i, g) 의 LFS 표본크기를 나타낸다. 주어진 Z_{ig} 에 대해 Z_{ig}^* 와 x_{ig} 의 독립성을 가정한다면 위의 식들은 다음과 같이 주어질 수 있다.

$$\begin{aligned} E(Z_{ig}^* | Z_{ig}) &= E\{E(Z_{ig}^* | Z_{ig}, x_{ig}) | x_{ig}\} \\ &= E(Z_{ig} | x_{ig}) \\ &= \pi_{ig} \end{aligned} \quad (3.96)$$

$$\begin{aligned} V(Z_{ig}^* | x_{ig}) &= E\left\{\frac{Z_{ig}(1 - Z_{ig})}{n_{ig}^*} | x_{ig}\right\} + V(Z_{ig} | x_{ig}) \\ &= \frac{\pi_{ig}(1 - \pi_{ig})}{n_{ig}^{**}} \end{aligned} \quad (3.97)$$

여기에서

$$n_{ig}^{**} = \frac{n_{ig}^{oo}}{1 + \frac{n_{ig}^{oo} - 1}{N_{ig}^*}}$$

이다. 일반적으로 n_{ig} 는 N_{ig} 에 비해 상대적으로 작은 값을 갖기 때문에 식 (3.96)와 (3.97)은 Z_{ig}^* 에 대한 일종의 근사 이항 로지스틱 모형을 정의하기 위하여 π_{ig} 에 대한 로지스틱 항과 결합될 수 있다. 이 모형은 표본 크기 $n_{ig}^o = \text{round}(n_{ig}^{**})$ 와 표본 실업자 수 $m_{ig} = \text{round}(n_{ig}^o \times Z_{ig}^*)$ 를 입력값으로 갖는 로지스틱 회귀 소프트웨어를 이용하여 실제 표본 자료에 적합될 수 있다. 여기에서 β 의 추정값과 $\text{Var}(\beta)$ 의 추정값 $v(\beta)$ 를 이끌어 낸다. 이때 Z_{ig} 의 추정량은 $\pi_{ig} = \text{antilogit}(x_{ig}^t \beta)$ 이 된다. 그러나 이 추정량은 불편성을 만족하지는 않는다. 따라서 편의를 보정한 형태의 추정량은 다음 (3.98)식과 같이 주어진다.

$$\pi_{ig} = \pi_{ig} \left\{ 1 - \frac{1}{2} (1 - \pi_{ig})(1 - 2\pi_{ig})(x_{ig}^t v(\beta) x_{ig}) \right\} \quad (3.98)$$

이때 소지역 g 에서 실업자 총계에 대한 추정량은 다음 (3.99)식과 같이 주어진다.

$$\theta_g = \sum_{i \in g} \alpha_{ig} N_{ig}^* \pi_{ig} \quad (3.99)$$

소지역 g 와 h 에 대한 모형기반 추정량들 간의 추정 공분산은 다음 (3.100)식을 통해 계산될 수 있다.

$$\widehat{\text{Cov}}(\theta_g, \theta_h) = \sum_{i \in g} \sum_{j \in h} \alpha_{ig} N_{ig}^* \pi_{ig} (1 - \pi_{ig})(x_{ig}^t v(\beta) x_{jh}) \pi_{jh} (1 - \pi_{jh}) N_{jh}^* \alpha_{jh} \quad (3.100)$$

위에서 언급된 합성추정 형태의 모형기반 추정량은 소지역들 간의 변동을 설명할 수 없는 문제점을 안고 있으며, 일반적으로 이러한 방법으로

추정된 추정오차는 과소 추정되는 경향이 있는 것으로 밝혀지고 있다. 이러한 문제점은 소지역에 대한 랜덤효과를 모형에 반영한 Multi-level 모형을 통해 어느 정도 해소할 수 있으며 이러한 연구가 현재 영국 통계국에서 진행되고 있다. 대상 모형은 $\text{logit}(\pi_{ig}) = x_{ig}^t \beta + u_g$ 와 같은 모형이다. 여기에서 지역 명시 변수 $\{u_g\}$ 는 평균이 0 이고 분산이 σ_u^2 인 확률변수로 가정된다. EBLUP 형태의 성분 추정값 $\pi_{ig} = \text{antilogit}(x_{ig}^t \beta + u_g)$ 에 기반을 둔 소지역 추정값들은 표본 크기가 큰 UA 및 LAD 지역에서는 LFS 추정값들과 유사하며, 표본 크기가 작은 UA 및 LAD 지역에서는 고정효과를 갖는 추정값들과 유사한 경향을 나타낸다. 이러한 추정량이 갖는 실제적인 문제는 추정량의 MSE 계산이 쉽지만은 않다는 데에 있다. 현재 영국 통계국에서는 하나의 절충안으로써 다음과 같은 분산 추정공식을 고려하고 있다.

$$\widehat{\text{Var}}(\theta_g) = \sum_{i \in g} \sum_{h \in g} \alpha_{ig} N_{ig}^* \pi_{ig} (1 - \pi_{ig}) \{ \sigma_u^2 + x_{ig}^t v(\beta) x_{ig} \} \quad (3.101) \\ \times \pi_{hg} (1 - \pi_{hg}) N_{hg}^* \alpha_{hg}$$

여기에서 σ_u^2 은 랜덤효과 모형 적합에서 추정되는 소지역 간의 추정분산을 나타낸다. 이상에서 고려된 방법론은 추정값의 정확도를 개선시키기는 하나, 모형에 랜덤효과를 포함시키는 문제와 EBLUP 형태의 추정량들의 MSE를 추정하는 방법 및 비 추정 방법 등은 여전히 해결되어야 할 문제점으로 남게 된다. 이러한 문제를 해결하기 위한 연구가 현재 영국 통계국 및 Southampton 대학 연구진들에 의해 진행되고 있다.

3.4 프랑스

노동력조사로부터 국가 수준의 노동 통계는 일년에 한 번씩 발표되고 있으며, 부차관심영역에서 실업 통계를 생산할 때 일정한 오차 범위 내로 기준을 충족하도록 하는 소지역 추정법에 관한 연구가 진행되고 있다.

프랑스의 표본 설계는 지역별 층화를 기준으로 하였으나 실업자 수 또는 실업률 등의 집적된 자료의 이용 시에는 무응답을 보정하여 국가 수준으로 통계를 작성한 후 성별-연령대별로 분할했으므로 지역 통계의 신뢰성에 대해서 유의해야 한다.

프랑스 통계청에서 취업과 실업에 관한 지역 통계의 생산을 위해서 노동력 조사와 센서스 뿐 만 아니라 실업 보험 신청자료 등 행정 업무 자료를 이용하는 체계를 발전시켜왔으나 소지역이나 세분화된 범주의 통계 작성은 미흡하다. 특히 취업 통계는 연말에 각 회사로부터 취업자 수에 대한 자료를 수집하고 있으나 전년말 기준으로 변화율에 대한 자료로 활용하여 국가 수준에서 비율에 대한 통계를 작성하고 있으며 실업자의 수에 대한 직접 통계를 작성하지는 않는다. 그래서 취업자 통계는 실업 보험과 센서스 자료 등이 핵심적인 역할을 하고 있다. 취업 통계는 해당 익년에 이용가능하다.

실업 통계는 분기별로 지역 통계와 국가 통계를 동시에 발표하며, 주로 이용되는 자료는 노동력 조사와 구직 등록자의 수이다. ILO 기준의 실업자와 실제 구직 등록자 및 노동력 조사에서 실업자와 차이를 반영하기 위해서 노동력 조사에서 추정된 ILO 실업자와 구직 등록자 수의 국가적 비율을 추정하여 활용하고 있다.

경제활동인구는 실업자와 취업자를 합해서 추정하지만 취업자는 직장을 갖고 있는 사람과 가사를 돌보는 사람을 합산해야 하므로 매 해마다 연말에 외삽법으로 조정하여 분기별 통계를 수정하여 시계열 통계로 관리한다.

취업자 통계에서 문제점은 취업자 통계를 작성 시 연말에 센서스 자료를 이용하고 있으나 센서스 간격이 너무 길어서 인구 변동과 상황 변화를 제대로 반영할 수 없다는 점이다 (센서스:1968, 1975, 1982, 1990, 1999). 특히, 지역 취업 통계 작성 시에는 더 심각해질 수 있다.

실업 통계 작성에서 문제점은 노동력 조사에서 국가 단위의 ILO 실업 통계는 작성할 수 있으나 지역 단위 또는 소지역 단위의 실업 통계는 생

산할 수 없다. 또한, 인구 사회학적 구성과 같은 구조적인 특성에서 각 지역이 국가와 동일하다는 전제에서 지역별 실업통계가 작성되며 표준오차는 계산하지 못하고 있다.

취업 통계 작성의 취약점을 보완하기 위해서 “ESTEL”이라는 프로젝트가 진행 중에 있으며 1996년 통계를 시험 중에 있다. 이 프로젝트의 특징은 매년 말을 기준으로 조사자료 및 행정자료를 기반으로 전국 및 지역 단위 취업 통계의 질과 신뢰도를 제고하는데 있다. 센서스 자료는 센서스 해에만 적용하고 다른 해에는 연말의 행정 업무와 조사 자료를 기준으로 한다는 것이 큰 특징이다.

핵심적인 과제는 경제 활동 관련 통계의 기준은 센서스 정보이므로 센서스 실시를 일정한 간격으로 실시하는 계획일 것이다. 매 5년마다 센서스에 준하는 대규모 통계 조사 계획을 추진하고 있다.

3.5 일 본

3.5.1 서 론

일본의 노동력조사는 매월 전국의 약 4만 세대, 10만 여명을 대상으로 실시하는 대규모 표본조사로써 노동력관련 특성치들인 취업자 수, 실업률 등은 이러한 월별 노동력조사를 통해 작성된다. 제한된 표본에서 추정결과의 정확도를 높이기 위해 표본설계에서는 국세조사구를 1차 추출단위로 하고, 가구를 2차 추출단위로 하는 층화 2단 추출법이 적용되며 이러한 층화 2단 추출법에 의해 전국의 11개 권역에 대해 조사객체를 선정하고 있다.

1차 추출단위의 추출은 가중치가 부여된 계통추출이고, 2차 추출단위의 추출은 등확률 계통추출이다. 또한 표본교체는 연동교체표본 설계를 따르며, 1차 추출단위인 국세조사구는 5년 마다 새로운 조사구들로 개편된다. 1차 추출단위의 추출틀은 조사구 명부이고, 조사구 명부는 그 특성에 의해서 층화되며 일정 순서에 의해 배열되어 있다. 2차 추출단위의 추출틀은 조사구 내의 가구명부가 이용된다.

월 노동력조사의 표본크기는 1차 추출단위가 약 2,900개 조사구이고, 2차 추출단위가 약 40,000개 가구로 구성된다. 노동력조사 결과에 대해서 시계열 분석을 시행할 경우 가장 많이 적용되는 방법은 전월과의 비교 및 전년 동월과의 비교이다. 일본 노동력조사의 표본설계는 시계열의 정도 향상을 위해 다음과 같은 특성을 갖는 연동교체표본 방식을 채택하고 있다. 첫째, 추출된 표본조사구는 4개월 간 계속해서 조사가 이루어지며 이후 매월 1/4씩 새로운 조사구로 교체된다. 즉, 추출된 표본조사구를 4개로 나누어 3/4이 전월과 같은 조사구, 1/4이 새로운 조사구로 대체되어 조사가 이루어진다. 둘째, 표본조사구는 4개월 간의 조사를 마친 후 8개월 간 조사에서 분리되어 익년 동기에 재 조사된다. 셋째, 표본조사구는 4개월 간 계속해서 조사를 시행하지만 조사구 내의 조사가구는 전반의 2개월과 후반의 2개월로 나누어 조사객체(조사가구)를 교체한다. 따라서 1차 추출단위로써 추출된 표본조사구는 4개월 간 계속해서 조사가 이루어지고, 이 기간이 끝나면 다른 표본조사구로 교체되지만 다음 해의 동기간에는 앞의 표본조사구에서 다시 조사가 이루어진다. 표본조사구의 교체는 일체히 이루어지지 않고, 전체를 A, B, C, D의 4개로 구분하고 매월 1구분씩 교체가 이루어진다. 2차 추출단위로써 추출된 가구는 2개월 간 계속해서 조사가 이루어지지만 표본조사구의 조사기간 4개월 중 전반의 2개월과 후반의 2개월로 가구를 교체한다. 따라서 추출된 가구에 거주하는 세대는 같은 가구에 거주하고 있으면 2개월 간 계속해서 조사가 이루어지고 또 다음 해의 같은 기간에 다시 2개월 간 계속해서 조사가 이루어진다.

일본 노동력조사의 표본설계 변천과정을 살펴보면 다음과 같이 요약될 수 있다. 일본 노동력조사는 매월 만 15세 이상의 노동력인구를 대상으로 1946년 9월에 시험적으로 개시되었다. 이때의 노동력조사는 1차 추출단위가 시, 군 단위, 2차 추출단위는 구, 동, 읍면 단위, 3차 추출단위는 세대 단위인 3단 추출법에 의해 추출된 약 15,000세대의 50,000여명의 고정표본에 대해 이루어졌다. 1948년 10월에는 2차 추출단위를 상주인구조사구로 조정하여 약 1,000조사구, 16,000여 세대의 56,000여명의 표본에 대해 조사

가 이루어졌으며, 표본조사구를 4개월 마다 일제히 교체하는 표본교체방식이 처음으로 실시되었다. 1950년 6월 노동력조사부터 2차 추출단위가 국세조사구로 개편되어 약 1,000개 조사구, 16,000여 세대의 51,000여명의 표본에 대해 조사가 이루어졌다. 1952년 11월에는 기존의 3단 추출법이 2단 추출법으로 변경되었다. 1차 추출단위는 1950년의 국세조사구, 2차 추출단위는 세대로 하여 추출된 약 1,000개 조사구, 11,000여 세대의 50,000여명의 표본에 대해 조사가 이루어졌고, 추출된 조사구는 3개월 간 계속 조사가 이루어진 후 매월 1/3의 조사구가 갱신되는 표본교체방식이 적용되었다. 1961년 10월부터 노동력조사 표본이 확대 개편되었다. 15세 이상의 노동력 인구를 대상으로 약 2,000개 조사구, 25,000여 세대의 70,000여명의 표본에 대해 조사가 이루어졌다. 추출된 표본조사구는 4개월 간 계속하여 조사가 이루어지고 매월 1/4의 조사구가 교체되는 표본교체방식이 채택되었고, 현재의 표본교체방식은 이 시기의 표본교체방식을 따르고 있다. 1982년 10월, 노동력조사에 대한 표본확대 개편이 시행되었다. 대영역을 기반으로 하는 기존의 표본설계 방식이 지역별 노동력 관련 통계를 생산하기 위한 체제로 확대 개편되었으며, 약 2,900개 조사구, 40,000여 세대의 100,000여명의 표본에 대해 노동력조사가 실시되었다. 현재의 노동력조사의 표본크기 및 표본교체방식은 이 시기의 표본설계에 기반을 두고 있다.

3.5.2 총화 및 추출단위 구성

일본의 노동력조사에서는 국세조사의 조사구를 1차 추출단위로 한다. 1개 조사구는 약 50세대로 구성되며 전국에 걸쳐 중복되지 않게 설정되어 있다. 95년 국세조사의 조사구 수는 전국적으로 약 88만개이다. 국세조사구를 추출단위로 사용하면 국세조사구 명부를 추출틀로 이용할 수 있고 그 밖에 국세조사의 조사결과를 이용해서 조사구를 그 특성에 의해 분류하는 일이 용이하게 될 수 있는 이점도 있다. 또한 조사구가 특별시, 광역시, 도 및 시, 구, 동/읍면의 행정구획 중에서 설정되어 있는 것도 조사

시 이점으로 작용하고 있다. 2차 추출단위는 표본조사구 내에 있는 주택과 건물의 각 호에서 1세대가 거주할 수 있도록 되어있는 건물 또는 건물의 일구획으로 정의된다. 이와 같이 정의된 2차 추출단위를 가구라고 한다. 2차 추출단위에 대한 결정규칙은 다음 <표3.39>와 같다.

<표3.39> 2차 추출단위의 결정규칙

건물의 종류	추출단위의 결정 규칙
단독주택	건물전체를 하나의 추출단위로 함.
연립주택, 아파트	하나의 세대가 사용하도록 만들어진 각각의 구획을 추출단위로 함.
기숙사 (하숙집, 식당, 수용 시설, 간이 숙박소 등을 포함)	기숙인 등의 각 거주실 및 가구주와 관리인 등의 세대 거주부분을 추출단위로 함. 식당, 수용시설 등에서 1실에 다수(약 10인 이상)의 거주자를 수용할수 있는 경우에는 1실을 5인 이하가 되도록 분할한 것을 하나의 추출단위로 함.
여관, 호텔	전 객실을 합쳐서 하나의 추출단위로 함
학교, 공장, 사무소 등	고용원 등의 세대가 거주할 수 있도록 되어있는 부분을 추출단위로 함.
병원, 요양소 등	입원환자의 각 병실 및 의사와 간호사 등의 거주부분을 각각 하나의 추출단위로 함. 병원 등의 1실에 다수의 입원환자가 들어가는 경우는 식당, 수용시설 등과 같이 분할한 것을 하나의 추출단위로 함.
주인집에서 숙식하는 사용인의 방이 3개 이상 있는 상점, 여관 등	사업주의 거주부분과는 별도로 주인집에서 사는 사용인의 각 방을 추출단위로 함. 사용인의 방이 3개 미만의 경우는 사업주의 세대에 포함함.
바깥채, 헛간 등의 부속 건물	주건물에 포함해서 하나의 추출단위로 함.
빈집, 공사 중의 건물 (사람이 전혀 살 수 없는 건물은 제외)	사람이 살고있는 건물에 준해서 추출단위로 함.

* 빈집, 공사 중의 건물도 추출단위로 하는 이유는 표본조사구를 4개월 간 계속해서 조사하는 것으로 하고 있기 때문에 2차 추출단위의 각 명부작성 시점에 거주자가 있지 않아도 조사시점에서는 거주자가 있을 가능성이 있기 때문이다.

일본 노동력조사의 표본설계에서는 1차 추출단위인 조사구의 층화에

중점을 두고 2차 추출단위인 가구에 대해서는 추출단위 명부 상의 배열을 고안하여 층화에 준한 효과를 얻고 있다. 더욱이 조사구의 층화는 전국을 11개 권역으로 나누고 각각의 권역에서 층화가 이루어지며, 중요도가 높은 통계항목인 산업별, 종사지위별 취업자 수에서 추정결과의 정도가 높도록 조사구의 산업별, 종사지위별 취업자 구성을 층화의 기준으로 이용하고 있다. 이밖에 기숙사, 합숙소, 병원, 요양소, 사회시설 등에 거주하고 있는 사람의 취업상태가 조사구의 특징을 결정하는 경우가 많기 때문에 조사구에 대한 주거형태도 층화기준에 더해진다. 일본 노동력조사의 표본 설계에서 층화에 대한 구체적인 방법은 다음과 같다.

○ 층부호 01

조사의 범위에서 제외되고 있는 주둔군 지역 및 추출대상에서 제외되고 있는 형무소, 구치소 등이 있는 구역, 자위대 구역 및 수면 조사구는 종합해서 하나의 층으로 한다.

○ 층부호 02, 03

사람이 살고 있지 않은 무인 조사구나 환산세대수가 15이하의 조사구에 대해서는 층화의 효과가 적기 때문에 산업별, 종사지위별 층화는 실시하지 않고 각각을 통합해서 하나의 층으로 한다. 환산세대수는 다음과 같이 산출한다.

$$\begin{aligned} \text{환산세대수} &= (\text{세대인원이 2인이상인 보통세대수}) \\ &+ \frac{1}{3} \{ (\text{세대인원이 1인인 보통세대수}) + (\text{준세대 인원}) \} \end{aligned}$$

보통세대는 주거와 생계를 함께하고 있는 사람들의 모임 또는 가구를 이루고 살고 있는 단독자를 말하며, 준세대는 자취, 하숙, 독신숙소의 단독자를 말한다. 위 식에서 세대인원이 1인인 보통세대수와 준세대 인원의 합을 3으로 나누는 것은 세대인원이 2인 이상인 일반세대의 15세 이상의 세대인원이 평균 3명을 가정한것에 기인한다.

○ 층부호 04

독신직원이 약 50인 이상 거주하는 기숙사, 합숙소가 있는 구역에 대

해서는 조사구 내에서 기숙사, 합숙소마다 단위구를 설정한다. 단위구마다 산업특성이 다르기 때문에 산업별 인구의 층 내 조사구 간의 분산을 줄이기 위해 층을 더욱 세분한다. 층의 세분화 과정에서 종사지위는 층화의 지표로 하지는 않는다. 학교의 독신학생이 약 50인 이상 거주하는 기숙사, 합숙소가 있는 조사구 및 사회시설이나 큰 병원이 있는 구역에 대해서는 시설의 종류에 따라 층을 설정한다.

○ 층부호 05~24

환산세대수가 16이상의 조사구는 조사구 내의 15세 이상 인구에 대한 산업별, 종사지위별 취업자 수의 비율에 의해 층화한다.

이상의 방법에 의해 설정된 조사구의 층화기준은 <표3.40>과 같다.

<표3.40> 조사구의 층화기준

분류부호		분 류 기 준
대분류	소분류	
01		후치번호가 5(형무소, 구치소 등이 있는 지역), 6(자위대 구역), 7(주둔군 지역), 9(수면조사수)인 조사구
02		인구가 0인 조사구
03		환산세대수가 15이하인 조사구
04		후치번호가 4(사회시설, 큰 병원이 있는 구역)인 조사구
		후치번호가 8(50인 이상의 단독자가 거주할 수 있는 기숙사, 합숙소 등이 있는 구역)인 조사구
		후치번호가 4와 7이외의 환산세대 중 점유할 수 있는 주택의 일반세대 비가 0.75이상인 조사구
	01	학생의 합숙소, 기숙사(단, 50인 이상의 세대)가 있는 표본단위구
	02	병원, 요양소(단, 50인 이상의 세대)가 있는 표본단위구
	03	사회시설(단, 50인 이상의 세대)이 있는 표본단위구
	04	후치번호가 4인 주택, 상기의 어디에도 속하지 않는 표본 단위구
	11	합숙소 등에 거주하는 광업의 취업자가 50인 이상의 표본 단위구
	12	광업의 세대비가 0.6이상의 표본단위구
	21	합숙소 등에 거주하는 건설업의 취업자가 50인 이상의

		표본단위구
	22	건설업의 세대비가 0.6이상의 표본단위구
	31	합숙소 등에 거주하는 제조업의 취업자가 50인 이상의 표본단위구
	32	제조업의 세대비가 0.6이상의 표본단위구
	41	합숙소 등에 거주하는 도소매업, 음식점의 취업자가 50인 이상의 표본단위구
	42	도소매업, 음식점의 세대비가 0.6이상의 표본단위구
	51	기숙사 등에 거주하는 금융·보험업, 금융업의 취업자가 50인 이상의 표본단위구
	52	광업의 세대비가 0.6이상의 표본단위구
	61	합숙소 등에 거주하는 전기, 가스, 열공급, 수도업, 운수 및 통신업의 취업자가 50인 이상의 표본단위구
	62	전기, 가스, 열공급, 수도업, 운수 및 통신업의 취업자의 세대비가 0.6이상의 표본단위구
	71	합숙소 등에 거주하는 서비스업의 취업자가 50인 이상의 표본단위구
	72	서비스업의 세대비가 0.6이상의 표본단위구
	81	합숙소 등에 거주하는 공무의 취업자가 50인 이상의 표본단위구
	82	공무의 취업자의 세대비가 0.6이상의 표본단위구
	91	후치번호가 8인 조사구가 있는 주택, 상기의 어디에도 속하지 않는 표본조사구
	92	급여주택에 살고 있는 일반세대수의 비가 0.75이상의 조사구가 있는 주택, 상기의 어디에도 속하지 않는 표본조사구
05		광업의 취업자의 비가 0.1이상의 조사구
06		어업의 취업자의 비가 0.2이상의 조사구
07		어업의 취업자의 비가 0.1이상-0.2미만의 조사구
08		건설업, 제조업의 업주의 비가 0.1이상의 조사구
09		도소매업, 음식점의 업주의 비가 0.1이상의 조사구
10		전기, 가스, 열공급, 수도업, 운수, 통신업, 금융, 보험업, 부동산업, 서비스업의 업주의 비가 0.1이상의 조사구
11		농림업의 취업자의 비가 0.3이상의 조사구
12		농림업의 취업자의 비가 0.1이상-0.3미만의 조사구
13		공무의 취업자의 비가 0.1이상의 조사구
14		금융, 보험업, 부동산업의 고용자의 비가 0.1이상의 조사구
15		제조업의 고용자의 비가 0.3이상의 조사구
16		건설업 고용자의 비가 0.1이상의 조사구

17		도소매업, 음식점의 고용자의 비가 0.2이상의 조사구
18		서비스업의 고용자의 비가 0.2이상의 조사구
19		전기, 가스, 열공급, 수도업, 운수, 통신업의 고용자의 비가 0.1이상의 조사구
20		제조업의 고용자의 비가 0.2이상-0.3미만의 조사구
21		제조업의 고용자의 비가 0.1이상-0.2미만의 조사구
22		도소매업, 음식점 고용자의 비가 0.1이상-0.2미만의 조사구
23		가스업의 고용자의 비가 0.1이상-0.2미만의 조사구
24		상기의 어디에도 속하지 않은 조사구

*단독세대는 사회시설, 큰 병원 및 약 50인 이상의 단독자가 거주하고 있는 기숙사, 합숙소 등을 말한다.

*총부호 04에 속하는 조사구에 대한 단위구의 설정

-후치번호가 4 또는 8의 조사구로 단독세대에 속하는 인원이 50인 이상이 되는 단독세대가 있는 경우는 단독세대별로 단위구를 설정한다.

-후치번호가 4, 8이외의 조사구에 대해서는 그 조사구 전체를 하나의 단위구로 한다.

*총부호 0411, 0421, 0431, 0441, 0451, 0461, 0471, 0481의 단위구는 후치번호가 8이다.

*총부호 0412, 0422, 0432, 0442, 0452, 0462, 0472, 0482의 단위구는 후치번호가 8이외이고 환산세대수 중에 차지하는 각 해당 산업 종사 세대주의 비가 0.6 이상의 단위구이다..

3.5.3 표본배분, 추출, 교체방식

일본 노동력조사에서는 조사구 추출 시 확률비례추출을 적용한다. 각 조사구 추출에 대한 가중치 산정은 다음공식을 적용한다.

$$\text{가중치} = \frac{\text{산출세대수}}{15}$$

총부호 01, 02의 조사구에 대해서는 가중치 1이 적용되며, 2차 추출단위의 추출율은 가중치의 역수를 사용한다.

노동력조사에서는 전국 추정결과를 산출하는 것을 주 목적으로 하며 10개 지역별 추정결과는 이를 기반으로 사분기마다 작성된다. 지역별 추정결과에 대한 목표정도를 만족할 수 있도록 다음과 같은 기준에 의해 조사구 수가 배분되고 있다.

○ 최소지역(시코쿠 지역)의 표본조사구 수

인구 규모가 가장 작은 시코쿠 지역의 표본조사구 수는 추정결과의 정도를 일정하게 유지할 수 있도록 표본조사구를 배분하며, 표본조사구의 추출은 8개 조의 부표본마다 행하기 때문에 부표본의 정수배로 조정하여 152개의 표본조사구를 배분한다.

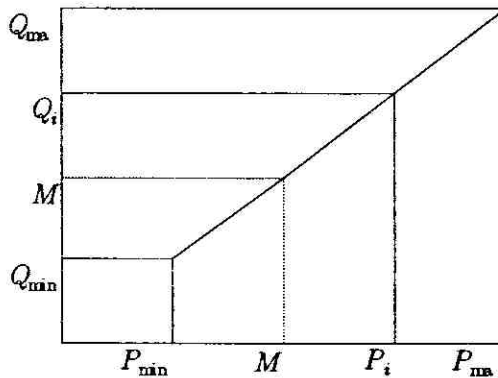
○ 오키나와 지역의 표본조사구 수

오키나와 지역의 도별 추정결과의 정도를 일정하게 유지할 수 있도록 부표본의 정수배인 144개 표본조사구를 배분한다.

○ 시코쿠 지역 이외지역에 대한 표본조사구 수

시코쿠 지역에 배분한 152개 조사구와 오키나와 지역에 배분한 144개 조사구를 제외한 나머지 2,584개 조사구는 선형변환법에 의해 인구비례적으로 각 지역별로 부표본의 정수배가 되도록 배분한다. 선형변환법은 인구비에 의한 배분값 P_i 를 다음의 일차식에 의해 재배분하는 방법이다.

$$Q_i = \frac{M - Q_{\min}}{M - P_{\min}} (P_i - P_{\min}) + Q_{\min}$$



여기에서 N 은 “전국의 표본조사구 수”, M 은 “1개 지역 당 평균 조사구 수(= $N/10$)”, P_i 는 “ i 지역의 비례배분값”, P_{\min} 은 “최소지역의 비례배분값”, Q_i 는 “변환 후 i 지역의 표본조사구 수”, Q_{\min} 은 “최소지역의 표본조사구 수”를 나타낸다.

지역별 표본조사구 수는 다음 <표3.41>과 같다. 각 지역별 표본조사구 수는 각 층의 조사구의 가중치 합계에 따라 비례배분하고 부표본의 정수 배로 조정해서 층별 표본조사구 수를 결정한다. 조사구의 가중치 합계가 현저하게 작은 층은 특성이 유사한 층과 합병한다. 단, 층부호가 02, 03, 0401, 0402, 0403, 0404의 각 층은 서로 상이한 속성을 갖기 때문에 합병을 시행하지 않고 각각 독립으로 표본조사구의 추출을 시행한다.

<표3.41> 지역별 표본조사구 수

지 역	특광역시 및 도	조사구 수
홋카이도	홋카이도	176
토오호쿠	아오모리, 이와테, 미야기, 아키타, 야마가타, 후쿠시마	232
미나미칸토	사이타마, 치바, 도쿄, 카나가와	576
기타칸도·코오신	이바라키, 토치기, 군마, 야마나시, 나가노	232
호쿠리쿠	니이가타, 토야마, 이시카와 후쿠이	168
도오카이	기후, 시즈오카, 아이치, 미에	304
킨키	시가, 교토, 오사카, 효고, 나라, 와카야마	400
츄우고쿠	돗토리, 시마네, 오키야마, 히로시마, 야마구치	208
시코쿠	토쿠시마, 카가와, 에히메, 코오치	152
큐슈	후쿠오카, 사가, 나가사키, 구마모토, 오이타, 미야자키, 가고시마	288
오키나와	오키나와	144
계		2,880

일본 노동력조사에서 표본조사구의 추출은 추출작업의 간편성, 정도의 향상 및 표본의 지역별 분산 등을 고려하여 계통추출법을 채택하고 있다. 계통추출법은 추출단위를 일정의 순서로 배열하고 이것에 추출용 일련번호를 붙여 추출기부호(계통추출의 출발점이 되는 부호)와 추출간격에 의해 추출용 일련번호를 순차적으로 추출하는 방법이다. 표본조사구는 각 지역별, 지역 내에서는 층별, 층 내에서는 8개 조의 부표본별로 독립적으

로 추출된다.

각 층 내에서 조사구의 배열은 “합병 전의 층”→“특·광역시 및 도”→“시·구·동/읍, 면”→“조사구 번호의 주 번호”→“조사구 번호의 단위구 번호”의 순으로 한다. 각 층별 조사구들에 대한 가중값의 누적합은 j 조사구의 가중값을 W_j 라 할 때 위의 조사구 배열 순서에 따라 $A_1 = W_1$, $A_2 = A_1 + W_2$, ..., $A_N = A_{N-1} + W_N$ (N 은 조사구 총수)와 같이 산출한다. 위식에서 A_j 를 “추출용 층내 누적부호”라 부른다. 층별 추출간격은 다음 식으로 계산한다.

$$i\text{층의 추출간격} = \frac{i\text{층 조사구의 가중값 합계}}{i\text{층의 표본조사구수}/8}$$

단, 층부호가 02, 0401, 0402, 0403, 0404인 6개의 각 층에 대해서는 “추출간격=조사구의 가중값 합계”로 추출간격이 결정된다.

<표3.41>에서 홑카이도 지역에 대한 층별 표본조사구 수 및 관련값들을 살펴보면 다음 <표3.42>와 같다.

<표3.42> 홑카이도 지역의 층별 표본조사구 등 일람표

합병전층부호	95년 국세조사		비례배분값	합병후층부호	표본조사구수	추출간격	추출기부호								교체조사구추출간격
	조사구수	가중값					AK	AL	BK	BL	CK	CL	DK	DL	
02	502	502	0.62	02	1	502					224			41	
03	3098	3098	3.83	03	3	3098			699				777	2493	258
04 01	121	336	0.42	04 01	1	336	127								28
04 02	255	930	1.15	04 02	1	930				640					77
04 03	356	920	1.14	04 03	1	920				61					76
04 04	117	243	0.30	04 04	1	243		234							20
06	732	2149	2.66												
07	461	1387	1.71												
12	1208	3458	4.27	06	16	6264	1069	1945	1963	2239	2771	3150	4002	5583	522
11	2440	5534	6.84												

04	31	14	38	0.05	04	31	16	5319	126	490	1567	1784	2723	2838	3926	4972	443
04	32	1	2	0.00													
15		150	475	0.59													
20		234	764	0.94													
21		2669	9265	11.45													
04	11	0	0	0.00													
04	12	0	0	0.00													
05		27	95	0.12													
04	21	1	2	0.00	04	21	32	6322	707	762	888	1377	2750	3417	4146	5534	526
04	22	0	0	0.00													
16		7407	25282	31.25													
08		2	6	0.01	04	41	8	10011	657	1430	1703	4802	5242	6950	7842	7862	834
04	41	1	5	0.01													
04	42	0	0	0.00													
17		2832	8445	10.44													
09		594	1561	1.93	22		48	6542	242	1137	1706	1987	4297	4331	5568	6483	545
22		10982	39256	48.53													
04	71	17	45	0.06	04	71	16	7425	108	1371	1968	3915	4407	4657	5072	5932	618
04	72	2	5	0.01													
18		4560	14801	18.3													
23		3191	10978	13.57	10		16	5577	149	620	857	1361	1474	2378	4429	4760	464
10		71	177	0.22													
04	51	1	2	0.00	04	51	8	4539	288	435	1518	2387	2530	2771	3389	4152	378
04	52	0	0	0.00													
14		369	1078	1.33													
04	61	5	12	0.01													
04	62	0	0	0.00													
19		1016	3447	4.26													
04	81	2	5	0.01													
04	82	1	2	0.00	04	81	8	8082	29	444	1477	1752	2923	5213	5703	6587	673
13		1570	5375	6.64													
04	91	48	106	0.13													
04	92	3	6	0.01													
24		863	2588	3.20													
계		45923	142380	176													

여기에서 22층에 대한 추출부호 산출과정을 세부적으로 살펴보면 다음

<표3.43>과 같다.

<표3.43> 홋카이도의 22층에서 추출부호의 산출

구 분	부 표 본								
	AK	AL	BK	BL	CK	CL	DK	DL	
가중값 합계	39256								
표본조사구수	6 (48÷8)								
추출간격(F)	6542.66 (39256÷8)								
추출기번호(G)	242	1137	1706	1987	4297	4331	5568	6483	
추출번호	$S_1 = G$	242	1137	1706	1987	4297	4331	5568	6483
	$S_2 = S_1 + F$	6785	7680	8249	8530	10840	10874	12111	13026
	$S_3 = S_2 + F$	13328	14223	14792	15073	17383	17417	18654	19569
	$S_4 = S_3 + F$	19871	20766	21335	21616	23926	23960	25197	26112
	$S_5 = S_4 + F$	26414	27309	27878	28159	30469	30503	31740	32655
	$S_6 = S_5 + F$	32957	33852	34421	34702	37012	37046	38282	39198

2차 추출단위인 가구조사 단위에 대한 추출은 추출단위 명부를 이용하여 계통추출한다. 이 경우 추출률은 조사구별로 정해져 있는 가중값의 역수를 이용한다.

노동력조사의 표본설계에서 표본교체방법을 살펴보기로 한다. 처음 추출한 표본조사구를 기준조사구라 하고, 소정의 조사기간을 마치고 교체되는 조사구를 교체조사구라 하자. 원칙적으로 기준조사구가 표본으로 선정되면 4개월 간 조사된 후 익년 동일 시기에 포함시키며 4개월 간 조사하고 표본에서 삭제한다. 교체조사구의 추출은 기준조사구의 추출용일련번호를 시작번호로 하고 교체조사구 추출간격을 이 번호에 더하는 계통추출법을 적용하며 구체적인 내용은 <표3.44>와 같다. 표본은 8개 조의 부표본으로 구성되어 있고, A, B, C, D의 구분은 조사를 새로 시작하는 달에 따라 붙여진 것이며, 후치번호 1은 금년에 새로 표본조사구에 포함된 조사구, 후치번호 2는 작년의 표본조사구로 포함되어 금년에 재조사한 후 내년에는 조사에서 삭제되는 조사구를 나타낸다.

<표3.44> 표본조사구 및 조사객체(가구)의 계속 상황

조별부호	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	
A-1(1년차) A-2(2년차)	[]		[]				전기 1월 2월		후기 3월 4월차		[]		
B-1(1년차) B-2(2년차)													
C-1(1년차) C-2(2년차)	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	
D-1(1년차) D-2(2년차)	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	

보다 구체적으로 각 조별 교체시기를 살펴보면 <표3.45>와 같다.

<표3.45> 조별 교체시기 현황

조별 부호		조사를 새로 시작하는 달
1년차 조사구	2년차 조사구	
A-1	A-2	1월 5월 9월
B-1	B-2	2월 6월 10월
C-1	C-2	3월 7월 11월
D-1	D-2	4월 8월 12월

A조에 속하는 조사구는 5월에 조사를 새로 시작해서 4개월 간 조사를 실시하고 새로운 A조의 조사구를 9월에 교체조사구로 추출하여 표본조사구로 포함시킨다. 그러나 5월에 조사를 시작한 조사구에 속하는 표본가구들은 5, 6월의 전반 2개월(전기) 조사분과 7,8월의 후반 2개월(후기) 조사분으로 분리 추출하여 조사한다. 즉 2차 추출단위인 표본가구들은 2개월 간 계속 조사한 후 1년 간 쉬었다가 익년의 같은 달에 조사된다. 하나의 기준조사구에 대응되는 교체조사구는 예비조사구를 포함하여 11개 조사구

를 추출하며 따라서 추출간격은 기준조사구 추출간격의 1/12로 한다. 2차 추출단위인 표본가구의 교체는 표본조사구 내에서 가구단위를 추출할 때 전기분과 후기분으로 나누어 별도의 독립적인 추출시작번호에 의해 계통추출한 표본가구들을 반씩 나누어 표본에서 삭제시키고 새로운 표본가구를 포함시킨다.

3.5.4 추정방법

표본조사 결과를 도출해 내는 가장 통상적인 방법은 각 조사결과에 추출률의 역수를 곱하여 합산한 후 추정값을 구하는 방법이다. 이와 같은 방법에 의해 구해진 추정값을 선형추정값이라 한다. 표본조사는 미지의 정보를 표본으로부터 획득하기 위해 시행된다. 이 경우 추정대상이 되는 목적항목과 연관성이 있는 보조정보가 있을 때 이것을 표본설계에 반영하면 조사결과의 정보량이 많아지고, 또한 결과의 정도도 향상될 수 있다. 일본 노동력조사에서는 목적항목에 대한 추정방법으로 보조정보를 이용한 다음과 같은 비추정식을 이용한다.

$$\text{목적항목의 비추정식} = \text{목적항목의 선형추정식} \times \frac{\text{보조항목의 참값(벤치마크)}}{\text{보조항목의 선형추정값}}$$

목적항목과 보조항목의 관련성은 목적항목의 종류에 따라 다르지만 양자의 상관성이 높을 때 추정결과의 정도는 좋아진다. 일본 노동력조사에서는 보조항목으로써 조사인구를 채택하고 있다. 노동력조사의 목적항목 및 보조항목의 선형추정값 X 는 다음 일차식에 의해 구해진다.

$$\hat{X} = \sum_{h=1}^{10} \hat{X}_h = \sum_{h=1}^{10} \sum_{i=1}^{L_h} \hat{X}_{hi} ,$$

단,

$$\begin{aligned} \hat{X}_{hi} &= \frac{1}{m_{hi}} \sum_{j=1}^{m_h} \frac{1}{P_{hij}} r_{hij} \left(\frac{1}{f_{hij}} \right) X_{hij} \\ &= \frac{1}{m_{hi}} \sum_{j=1}^{m_h} \frac{W_{hi}}{w_{hi}} r_{hij} f_{hij} X_{hij} \end{aligned}$$

$$= F_{hi} \sum_{j=1}^{m_{hi}} r_{hij} X_{hij}$$

여기에서

h = 지역 번호(1, 2, ..., 10),

i = 층 번호(1, 2, ..., L_h),

j = 표본조사구 번호(1, 2, ..., m_{hi}),

\hat{X} = 속성 X 의 선형추정값,

\hat{X}_h = h 지역에 대한 속성 X 의 선형추정값,

L_h = h 지역의 층의 수,

\hat{X}_{hi} = h 지역의 i 층에 대한 속성 X 의 선형추정값,

m_{hi} = h 지역에서 i 층의 표본조사구 수,

W_{hi} = h 지역의 i 층에 대한 모든 조사구의 가중값 합계,

F_{hi} = h 지역의 i 층에 대한 선형추정 승률(= W_{hi}/m_{hi}),

X_{hij} = h 지역의 i 층에서 j 표본조사구에 대한 속성 X 의 조사인구,

P_{hij} = h 지역의 i 층에서 j 표본조사구의 추출확률,

w_{hij} = h 지역의 i 층에서 j 표본조사구의 가중값,

r_{hij} = h 지역의 i 층에서 j 표본조사구의 보정률(조사구 세부 분할에 따른 변경 추출률),

f_{hij} = h 지역의 i 층에서 j 표본조사구의 조사구 내 추출률의 역수

($f_{hij} = w_{hij}$).

일본 노동력조사에서 목적항목에 대한 추정값은 다음 그림과 같은 과정을 통해 산출된다.

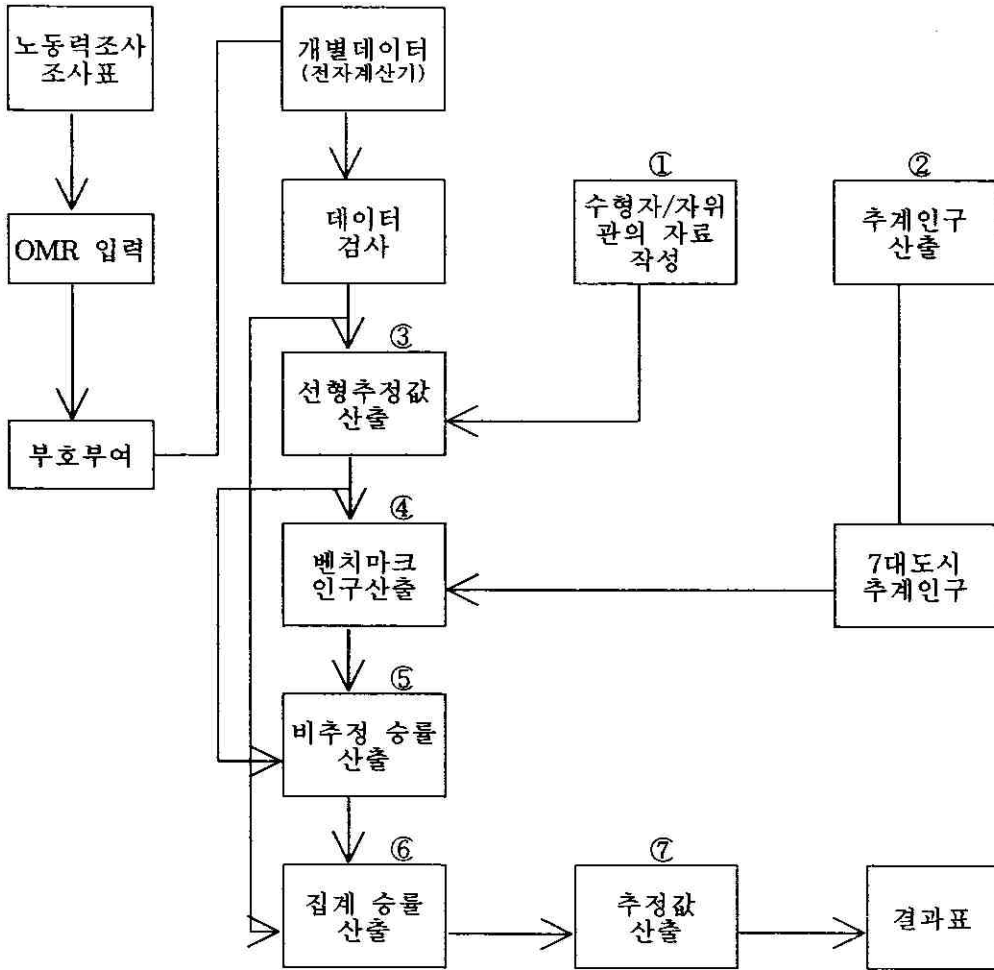


그림3.1 추정값 산출 순서

① 수형자 및 자위관의 자료 작성

노동력조사에서는 총화 2단추출법에 의해 조사가구가 추출된다. 그러나 형무소 등의 교정시설 수용자(수형자) 및 자위대의 당사 내 거주자(자위관)에 대해서는 조사원에 의한 조사가 곤란하므로 법무성 및 방위청으로부터 자료를 얻어 통계청에서 직접 직접하고 있다.

② 추계인구의 산출

국세조사에 의한 인구를 기준인구로 하고 매월 인구동태 통계의 출생아 수, 사망자 수, 출입국관리 통계의 출국자 수, 입국자 수의 통계 숫자를 가감하여 매월 말일 현재 전국의 남녀, 연령계급별 추계인구를 산출하고 있다. 추계인구의 산출식은 다음과 같다.

$$P = P_0 - P_d + P_m + P_s$$

여기에서 P_0 는 “전월말 현재 추계인구”, P_d 는 “ P_0 인구 중 이 달에 사망한 인구”, P_m 은 “금월 중 연령계급의 이동에 의한 증가 수”, P_s 는 “금월 중 입국자수에 출국자 수를 감한 인구”를 나타낸다.

③ 인구의 선형추정값 산출

각 조사가구에는 조사가구가 속한 조사구의 선형추정 승률이 부여되어 있다. 각 조사가구의 선형추정 승률을 남녀별(2범주), 지역별(토쿄, 요코하마, 나고야, 교토, 오오사카, 고베, 기타큐슈의 7대도시와 7대도시 이외의 지역의 2범주), 연령계급별(14범주)의 각 범주에서 합산하고 각 범주의 인구 선형추정값을 산출한다.

④ 벤치마크 인구 산출

앞의 ②에서 산출된 추계인구는 전국의 남녀(2범주)-연령계급(14범주)별 인구와 남녀(2범주)-지역(7대도시와 7대도시 이외 지역의 2범주)별 인구의 총 32개 범주에 대해서 작성된다. 벤치마크 인구는 32개 범주의 추계인구와 ③에 의해 산출된 선형추정값을 이용하여 지역별-남녀별-연령계급별의 총 56개 범주에 대해 산출된다.

다음 <표3.46>를 참고하여 구체적인 벤치마크 인구 산정에 대해 설명하기로 한다. 우선 <표3.46>에서 ○에는 ②의 추계인구, △에는 ③에서 산출된 선형추정값을 기입한다. △의 숫자의 세로합(Ⅲ)이 지역별 추계인구(Ⅰ)와 일치하지 않을 경우에는 (Ⅳ)의 보정 승률을 산출하여 △의 각 숫자에 곱해서 종계를 추계인구와 일치시킨다. 같은 방법으로 △의 보정숫자의 가로합이 추계인구와 같도록 보정한다. 이와 같은 방법을 각 범주에 대해 순차적으로 계속 적용하여 가로와 세로행렬에 대한 보정을 시

도하면 △의 숫자는 점점 안정되고, 가로와 세로의 관계에 모순이 없는 숫자로 수렴하게 된다. 최종적으로 얻어진 산출값을 56범주에 대한 벤치마크인구로 이용한다.

<표3.46> 56개 범주에 대한 벤치마크 인구의 산출표(남녀별)

구 분		(i)	(ii)	(iii)	(iv)	(v)
		전국	7대도시	7대도시 이외지역	총계	작수회계 보정승률 (i)÷(iv)
(I)	총수	○	○	○	-	-
(II)	0-14세	○	△	△		
	15-19세	○	△	△		
	20-24세	○	△	△		
	25-29세	○	△	△		
	30-34세	○	△	△		
	35-39세	○	△	△		
	40-44세	○	△	△		
	45-49세	○	△	△		
	50-54세	○	△	△		
	55-59세	○	△	△		
	60-64세	○	△	△		
	65-69세	○	△	△		
	70-74세	○	△	△		
75세이상	○	△	△			
(III)	총계	-				-
(IV)	홀수회계 보정승률 (I)÷(III)	-			-	-

⑤ 비추정 승률의 산출

남녀별, 지역별, 연령계급별로 다음 식에 의해 비추정을 위한 승률을 산출한다.

$$\text{비추정 승률} = \frac{\text{벤치마크인구}}{\text{선형추정값}}$$

참고로 일본 노동력조사에서 1998년 1월부터 12월까지의 15세 이상 비추정 승률의 평균값은 1.108이다.

⑥ 집계 승률의 산출

비추정값은 목적항목의 56개 범주의 선형추정값에 비추정 승률을 곱하여 구할 수 있다. 그러나 계산의 편의 상 각 조사가구마다 집계 승률을 산출해서 계산하고 있다.

각 조사가구의 집계 승률 = 조사가구가 속한 조사구의 선형추정 승률
× 조사가구가 속한 연령계급, 남녀, 지역의 비추정 승률

⑦ 추정값의 산출

위 ⑥의 “각 조사가구의 집계 승률”을 각 결과표의 목적항목마다 합산해서 구한 결과를 산출한다. 지역별 목적항목에 대한 추정결과는 월별 추정값들의 3개월 분을 평균하여 전국결과와 일치하도록 보정하여 산출한다.

3.5.5 오차관리

일본 노동력조사에서 표본오차의 계산은 조사결과를 사용하여 계산되기 때문에 표본오차에 비표본오차도 일부 포함되어 계산된다. 모든 표본을 랜덤하게 8등분하여 얻어지는 각 부표본들은 전국에 대한 랜덤표본으로 간주되기 때문에 각 부표본에서도 동일 항목의 추정을 시행할 수 있다. 전국의 취업자 수에 대해서 표본 전체에서 추정한 결과를 \hat{X} , 각 부표본으로부터 추정한 결과를 $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_8$ 이라 하면 부표본에 의한 추정값 \hat{X}_i 는 전체 표본에서 추정된 결과와 대체로 비슷한 수치를 가질 것으로 기대되지만 \hat{X}_i 의 표본오차가 클 경우 흐트러짐의 정도가 크게 나타난다. 이러한 흐트러짐의 정도는 8개의 부표본에 의한 평균제곱오차

$$\frac{1}{7} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2$$

로 계산된다.

표본이론에 의하면 추정값의 분산은 표본의 크기에 역비례한다. 따라서 각 부표본에 의한 추정값의 분산은 전체 표본의 분할 수에 역비례하게

된다. 이것을 8개의 부표본별 추정값과 표본 전체에 의한 추정값과의 관계에 적용시키면, 표본 전체는 하나의 부표본의 8배가 되므로 표본 전체에 의한 추정값의 평균제곱오차는 각 부표본에 의한 추정값의 분산의 1/8이 된다. 즉 \hat{X} 의 평균제곱오차는

$$\left(\frac{1}{7} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2 \right) / 8 = \frac{1}{56} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2$$

로 계산되며, 표준오차는 \hat{X} 의 평균제곱오차의 제곱근으로 구해진다. 한편 변동계수 CV값은

$$\frac{\sqrt{\frac{1}{56} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2}}{\hat{X}}$$

로 계산된다.

1998년 1월부터 12월까지의 일본 노동력조사의 주요항목별 추정값, 표준오차, CV값들에 대한 12개월 분의 단순 평균값들이 다음 <표3.47>에 주어졌다.

<표3.47> 주요항목별 추정결과

주요항목	추정값 (만명)	표준오차 (만명)	CV (%)	
노동력 인구	6793	27.2	0.4	
산업별 취업자	전산업	6514	26.1	0.4
	농림업	317	10.5	3.3
	비농림업	6197	24.8	0.4
	어업	27	1.4	5.3
	광업	6	0.3	5.7
	건설업	662	13.2	2.0
	제조업	1382	16.6	1.2
	전기, 가스, 열공급, 수도업, 운수 및 통신업	442	9.3	2.1
	도소매업, 음식점, 금융, 보험업, 부동산업	1741	20.9	1.2
	서비스업	1685	16.9	1.0

	공무		217	5.9	2.7
농림업, 비농림업종 상의 지위별 취업자	전산업	자영업주	761	12.2	1.6
		가족종업자	367	9.5	2.6
		고용자	5368	21.5	0.4
	농림업	자영업주	156	5.0	3.2
		가족종업자	128	5.9	4.6
		고용자	34	2.3	6.7
	비농림업	자영업주	605	9.7	1.6
		가족종업자	240	7.9	3.3
		고용자	5334	21.3	0.4
완전 실업자			279	7.0	2.5
비노동인구			3924	27.5	0.7

1998년 1월부터 12월까지의 항목별 추정자료를 이용하여 추정값과 CV값을 곡선 모형에 적합시켜 추정값의 크기에 따른 CV값을 산정한 결과가 다음 <표3.48>에 주어졌다. 산정결과는 추정값의 규모에 따른 CV값의 목표허용정도를 결정하는 기준으로 참고한다.

<표3.48> 추정값의 크기별 CV값

(i) 전국 연평균 추정값의 CV값

추정값의 크기 (만명)	5000	3000	2000	1000	700	500	300	200	100	70	50	30	20	10
표준오차 (만명)	10.8	8.3	6.7	4.7	3.9	3.3	2.5	2.0	1.4	1.2	1.0	0.8	0.6	0.4
CV (%)	0.2	0.3	0.3	0.5	0.6	0.7	0.8	1.0	1.4	1.7	2.0	2.5	3.1	4.3

(ii) 월별 추정값의 CV값

추정값의 크기 (만명)	5000	3000	2000	1000	700	500	300	200	100	70	50	30	20	10
표준오차 (만명)	19.7	15.4	12.6	9.0	7.6	6.5	5.1	4.2	3.0	2.5	2.1	1.7	1.4	1.0
CV (%)	0.4	0.5	0.6	0.9	1.1	1.3	1.7	2.1	3.0	3.6	4.3	5.5	6.8	9.8

(iii) 10개 지역별 연평균 추정값의 CV값(단위:%)

지역	추정값의 크기(만명)													
	2000	1000	500	300	200	100	50	30	20	10	5	3	2	1
홋카이도				0.80	1.00	1.40	2.00	2.50	3.10	4.30	6.00	7.70	9.30	1.30
토오호쿠			0.60	0.80	1.00	1.40	1.90	2.50	3.00	4.10	5.80	7.40	8.90	1.25
미나미칸토	0.40	0.60	0.80	1.10	1.30	1.80	2.50	3.20	3.90	5.40	7.50	9.50	11.5	1.60
기타칸토· 코오신			0.70	0.90	1.10	1.50	2.10	2.70	3.20	4.40	6.10	7.70	9.30	1.29
호쿠리쿠				0.70	0.80	1.10	1.60	2.00	2.40	3.40	4.80	6.10	7.40	1.06
토오카이		0.40	0.60	0.80	1.00	1.30	1.90	2.50	3.00	4.30	6.00	7.80	9.60	1.35
킨키		0.60	0.80	1.00	1.20	1.60	2.30	2.90	3.50	4.90	6.70	8.60	10.4	1.44
츄우고쿠			0.70	0.90	1.00	1.50	2.00	2.60	3.10	4.30	5.90	7.50	9.10	1.26
시코쿠				0.60	0.80	1.10	1.50	1.90	2.40	3.30	4.60	6.00	7.30	1.02
큐슈		0.60	0.80	1.00	1.20	1.60	2.20	2.80	3.40	4.60	6.30	8.00	9.60	1.31

(iv) 10개 지역별 사분기 평균 추정값의 CV값

지역	추정값의 크기(만명)													
	2000	1000	500	300	200	100	50	30	20	10	5	3	2	1
홋카이도				1.40	1.70	2.30	3.30	4.20	5.20	7.30	10.3	13.3	16.2	22.8
토오호쿠			1.10	1.40	1.70	2.40	3.40	4.40	5.30	7.50	10.5	13.4	16.4	23.0
미나미칸토	0.70	1.00	1.40	1.80	2.20	3.00	4.30	5.40	6.60	9.30	13.0	16.7	20.3	28.4
기타칸토· 코오신			1.10	1.40	1.80	2.50	3.40	4.40	5.40	7.50	10.5	13.5	16.4	23.0
호쿠리쿠				1.20	1.40	2.00	2.80	3.60	4.40	6.20	8.70	11.2	13.7	19.4
토오카이		0.70	1.10	1.40	1.70	2.40	3.40	4.40	5.40	7.60	10.8	14.0	17.1	24.3
킨키		1.00	1.30	1.70	2.10	2.90	4.00	5.20	6.30	8.80	12.3	15.7	19.0	26.6
츄우고쿠			1.10	1.40	1.70	2.40	3.30	4.30	5.20	7.40	10.4	13.3	16.3	22.9
시코쿠				1.10	1.40	1.90	2.70	3.50	4.30	6.10	8.60	11.0	13.5	19.0
큐슈		0.90	1.20	1.60	1.90	2.60	3.70	4.70	5.70	7.90	11.0	14.0	16.9	23.6

제4장 소지역 추정법

4.1 개 요

지리적으로 구분된 영역 또는 좀 더 일반적으로 임의의 부분모집단이 소지역으로 간주될 수 있다. 지리적으로 구분된 도 단위가 소지역이 될 수도 있고, 도 단위 내의 시군구 단위가 소지역이 될 수도 있다. 통상적으로 소지역에 대한 표본의 크기는 소지역의 크기에 따라 증가하는 경향이 있으나, 대영역을 기준으로 작성된 표본설계에서 소지역을 추정할 경우에는 반드시 이러한 경향을 보이는 것은 아니다. 소지역에 대한 추정값들은 정부의 지역별 예산 배분 등의 정책 입안 시에 절대적인 참고자료로 활용될 수 있기 때문에 최근에는 신뢰할 수 있는 소지역 통계 작성에 많은 관심들을 보이고 있다.

5년 또는 10년 마다 실시되는 센서스 자료, 실업자의 구직 등록 자료나 실직 보험 자료와 같은 행정보고 자료, 다양한 사회적 관심에 의해 실시되는 표본조사 자료 등이 소지역 추정을 위해 활용된다. 그러나 많은 경우 표본조사에 의해 확보된 직접 추정값들은 대영역에 대해서는 신뢰할 만한 수준이나, 소지역에서는 확보된 표본이 작을 경우가 대부분이므로 신뢰할 만한 정확도를 기대하기는 어려운 실정이다. 이러한 이유 때문에 유사한 특성을 갖는 인근 지역으로부터 정보를 취득하여 소지역의 추정의 정확도를 높일 수 있는 간접 추정방법을 생각하게 되는데, 이러한 추정방법이 소지역 추정에 활용된다.

간접 추정량은 최근의 센서스 자료라든가 행정 보고 자료와 같은 보조 자료를 소지역들과 연계할 수 있는 암시모형(Implicit Model) 또는 명시모형(Explicit Model)을 설정하여 추정된다. 암시모형 하에서 추정되는 간접 추정량으로 합성 추정량(Synthetic Estimator)과 복합 추정량(Composite Estimator)을 들 수 있고, 모형에 근거한 추정량으로는 경험적 베이스(EB) 추정량, 계층적 베이스(HB) 추정량 등을 들 수 있다.

이번 장에서는 소지역 추정기법에 관한 일반적인 이론들을 요약 정리한다. 대표적인 간접 추정법인 인구통계학적 방법, 합성 추정법과 복합 추

정법을 살펴보고, 모형 기반 추정법으로는 *EB* 추정법과 *HB* 추정법을 소개한다.

4.2 인구통계학적 방법(Demographic Method)

미국의 경우에서와 같이 10년 주기로 센서스를 할 경우, 지방 도시나 county의 중간 해당 년도의 인구를 추정하기 위해서 사용하는 추정법으로 센서스 자료와 인구수에 관련된 장후 변수(출생자수, 사망자수, 주택수, 등록된 학생 수 등)의 변동을 분석하여 얻은 예측값을 결합하는 추정법을 인구 통계학적 방법(Demographic Method)이라 말한다.

4.2.1 생멸률법(Vital Rates Method: VR Method)

*VR*법은 출생과 사망에 관련된 자료를 이용하여 인구의 변동을보다는 장후 변수의 영향만을 분석하여 활용한다. 가장 최근에 센서스를 실시한 해를 기준 년도로 하고, 기준해로부터 t 년 후에 소지역의 인구 수를 추정한다고 하자. 이 때 전제 조건은 추정 대상인 소지역을 포함하는 대지역의 특성과 소지역의 특성이 동일하다는 것이며, 전제 조건에서 많이 벗어나는 경우에는 추정량의 편향이 커져서 신뢰도가 낮아진다.

t 년 후의 소지역의 출생률과 사망률을 γ_{bt} 와 γ_{dt} 로 표현하고 대영역의 출생률과 사망률을 R_{bt} 와 R_{dt} 라 나타내면 다음과 같은 관계가 주어진다.

$$\gamma_{bt} = \gamma_{bo} \left(\frac{R_{bt}}{R_{bo}} \right), \quad \gamma_{dt} = \gamma_{do} \left(\frac{R_{dt}}{R_{do}} \right), \quad (4.1)$$

여기에서 γ_{bo} 와 γ_{do} 는 기준 해의 소지역의 출생률과 사망률이고 R_{bo} 와 R_{do} 는 기준 해의 대지역의 출생률과 사망률을 의미한다.

센서스를 실시한 기준해로부터 t 년 후의 인구 수는 다음 식에 의해서 추정할 수 있다.

$$p_t = \frac{1}{2} \left(\frac{b_t}{\gamma_{bt}} + \frac{d_t}{\gamma_{dt}} \right). \quad (4.2)$$

단, b_t 와 d_t 는 소지역의 t 년 후의 출생자수와 사망자수를 뜻한다.

4.2.2 성분법(Component Method)

성분법은 출생과 사망 인구수 및 유입, 유출 인구에 관한 자료를 이용하여 소지역의 인구 수를 추정하기 위해 고안된 방법이다. 센서스를 실시한 기준해로부터 t 년 동안의 출생 인구, 사망인구 및 총 이주인구를 각각 $b_{0,t}$, $d_{0,t}$ 와 $m_{0,t}$ 로 나타냈을 때 t 년 후의 인구수는 다음식에 의해 추정한다.

$$p_t = p_0 + b_{0,t} - d_{0,t} + m_{0,t}, \quad (4.3)$$

여기에서 $m_{0,t} = i_{0,t} - e_{0,t} + n_{0,t}$ 로 주어지며, $i_{0,t}$ 는 유입인구, $e_{0,t}$ 는 유출인구, $n_{0,t}$ 는 주 간의 총 이주인구를 나타내며 행정보고자료에 의해 주어진다.

4.2.3 회귀 징후법(Regression Symptomatic Procedures)

회귀 징후법은 다중선형회귀모형을 이용하여 소지역의 인구를 추정하는 방법으로써 징후변수들을 독립변수로 선택하여 소지역 추정에 이용한다. 비 상관계수(ratio correlation), 차분 상관계수(difference correlation), 표본 회귀법(sample regression method) 등은 이러한 회귀징후법의 일종이다. 여기에서는 다른 두 방법보다는 비교적 자주 사용되고 있는 표본 회귀법을 설명하기로 한다. 먼저 종속변수와 독립변수를 다음과 같이 정의하자.

$$Y_i = (p_{it}/P_t)/(p_{i0}/P_0) = \text{소지역 } i \text{의 인구비 변화량,}$$

$$x_{ij} = (s_{ijt}/S_{jt})/(s_{j0}/S_{j0}) = \text{소지역 } i \text{에 대한 } j \text{번째 징후변수 } s_j \text{의 변화량,}$$

여기에서 P_t , P_0 , S_{jt} 와 S_{j0} 는 소지역 i 를 포함하는 대지역에서의 값들이고, x_{ij} 는 행정자료로부터 얻는다($j=1, 2, \dots, p$).

회귀 표본법은 종속변수 Y_i 가 징후변수 $x_{i1}, x_{i2}, \dots, x_{ip}$ 의 일차결합으로

로 표현될 수 있다는 것을 가정하며, 이때 Y_i 의 값은 조사된 직접추정값 \hat{Y}_i 을 이용하여 m 개의 소지역 중 k 개의 소지역에 대하여 선형회귀식을 적합시켜 회귀계수들을 추정한 후, Y_i 의 추정값으로 다음의 표본회귀 추정량을 이용한다.

$$\tilde{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}, \quad i = 1, 2, \dots, m \quad (4.4)$$

소지역 i 에 대한 인구수는 (4.4)식의 표본 회귀추정량을 이용하여 다음 식으로 추정한다.

$$\bar{p}_i = \tilde{Y}_i \left(\frac{p_0}{P_0} \right) \hat{P}_i, \quad i = 1, 2, \dots, m \quad (4.5)$$

표본 회귀추정량은 표본으로부터 직접 추정된 값이 아니라 다중 선형회귀를 거쳐 얻어진 보정된 추정량이며, 표본 회귀법은 이를 이용하여 소지역의 인구를 추정하는 방법이다. 그러나 이러한 방법은 추후 논의될 모형에 근거한 소지역 추정보다는 효율성이 상당히 떨어지는 것으로 밝혀지고 있다.

4.3 합성 추정법(Synthetic Estimation)

추정하고자 하는 소지역과 특성이 유사한 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 필요한 때에는 대영역과 소지역의 구조적 특성이 같다는 조건하에서 소지역의 연구변수에 대한 추정값을 구할 수 있는데, 이때 대영역의 분할은 지리적인 분할보다는 연령대별 또는 교육정도별과 같은 특성에 따른 분할을 말한다.

대영역을 I 개 소지역으로 분할하며 또한 대영역을 특성 기준에 따라 J 개의 범주로 분류한다면 i 소지역의 추정값은 다음 식으로 구할 수 있다.

$$\hat{Y}_{i.} = \sum_j^J p_{ij} \hat{Y}_{.j} \quad (4.6)$$

단, p_{ij} 는 소지역 i 의 j 범주에 대한 가중값이며 센서스나 행정자료에서 구해진다. $\hat{Y}_{.j}$ 는 대영역에서 j 범주에 대한 표본 추정값이다. 단, 대영역의 표본의 수는 충분하게 많아서 신뢰성 있는 추정값을 구할 수 있다고 가정한다. i 소지역의 실업자 추정에 관한 경우를 생각해 보자.

Y_{ij} : i 소지역의 j 범주(연령대별 또는 교육정도별)의 실업자수,

X_{ij} : i 소지역의 j 범주(연령대별 또는 교육정도별)의 경제활동인구,

$Y_{.j} = \sum_i Y_{ij}$: 대영역에서 j 범주의 실업자 수,

$Y_{i.} = \sum_j Y_{ij}$: i 소지역의 실업자 수.

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

$$\hat{Y}_{i.}^s = \sum_j \left(\frac{X_{ij}}{X_{.j}} \right) \hat{Y}_{.j} \quad (4.7)$$

만약 $\hat{Y}_{.j}$ 가 비추정량의 형식을 갖는다면,

$$\hat{Y}_{.j} = \left(\frac{\hat{Y}_{.j}}{\hat{X}_{.j}} \right) X_{.j}$$

로 나타낼 수 있으므로 (4.7)식은 다음과 같이 표현될 수 있다.

$$\hat{Y}_{i.}^s = \sum_j X_{ij} \left(\frac{\hat{Y}_{.j}}{\hat{X}_{.j}} \right) = \sum_j \left(\frac{X_{ij}}{\hat{X}_{.j}} \right) \hat{Y}_{.j} \quad (4.8)$$

여기에서 $\hat{Y}_{i.}^s$ 가 불편추정량이 되기 위해서는 $\frac{Y_{.j}}{X_{.j}} = \frac{Y_{ij}}{X_{ij}}$ 를 만족해야 하고, 이를 만족하지 못할 경우에는 편향추정량이 되며, 이때 $\hat{Y}_{i.}^s$ 의 편향의

크기는 $B(\hat{Y}_{i.}^s) = E(\hat{Y}_{i.}^s - Y_{i.})$ 이다. 즉,

$$B(\hat{Y}_{i.}^s) = \sum_j X_{ij} \left(\frac{Y_{.j}}{X_{.j}} - \frac{Y_{ij}}{X_{ij}} \right).$$

$\hat{Y}_{i.}^s$ 의 평균제곱오차 $MSE(\hat{Y}_{i.}^s)$ 의 근사적 불편추정량은 다음과 같이 주어질 수 있다.

$$\widehat{MSE}(\hat{Y}_{i.}^s) = (\hat{Y}_{i.}^s - \hat{Y}_{i.})^2 - \widehat{Var}(\hat{Y}_{i.}) \quad (4.9)$$

4.4 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.

$$\hat{Y}_{i.}^C = w_i \hat{Y}_{i.} + (1 - w_i) \hat{Y}_{i.}^s \quad (4.10)$$

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

먼저 평균제곱오차 $MSE(\hat{Y}_{i.}^C)$ 를 최소화하는 w_i 는 아래와 같다.

$$w_{i(opt)} = \frac{MSE(\hat{Y}_{i.}^s)}{MSE(\hat{Y}_{i.}^s) + Var(\hat{Y}_{i.})} \quad (4.11)$$

최적 가중값 $w_{i(opt)}$ 의 추정값은 다음 식으로 계산될 수 있다.

$$\hat{w}_{i(opt)} = \frac{mse(\hat{Y}_{i.}^s)}{(\hat{Y}_{i.}^s - \hat{Y}_{i.})^2} \quad (4.12)$$

모든 소지역에 공통 가중값을 부여하는 방법으로써 초기 공통 가중값 w 를 이용하여 $MSE(\hat{Y}_{i.}^s)$ 들의 평균을 최소화하는 가중값은 아래와 같다.

$$\hat{w}_{(opt)} = 1 - \frac{\sum_i \widehat{Var}(\hat{Y}_{i..})}{\sum_i (\hat{Y}_{i..}^s - \hat{Y}_{i..})^2} \quad (4.13)$$

각 소지역에 배정된 표본 크기에 의존하는 가중값은 다음과 같이 계산된다.

$$w_i(\delta) = \begin{cases} 1, & \hat{N}_i \geq \delta N_i \\ \frac{\hat{N}_i}{\delta N_i}, & otherwise \end{cases} \quad (4.14)$$

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 대영역을 소지역으로 분할하여 각 소지역의 추정값을 추정하므로 소지역의 추정값의 합계는 대영역의 추정값과 같아야 할 것이다. 왜냐하면 매월 정부기관에서 발표하는 광역시와 도의 실업자 수와 해당 소지역의 추정값의 합계가 같도록 조정하지 않으면 서로 상이한 통계수치로 인하여 혼란을 줄 수 있기 때문에 한 가지 통계수치가 되도록 조정된 추정량을 계산해야 할 것이다. 각 소지역의 추정량을 생명률법, 합성추정법 또는 복합 추정법 등의 어느 한 방법으로 계산한 것으로 간주할 때 조정된 소지역 추정량은 다음과 같다.

$$\hat{Y}_{i..}^A = \left(\frac{\hat{Y}_{i..}}{\sum_i \hat{Y}_{i..}} \right) \hat{Y} \quad (4.15)$$

단, \hat{Y} 는 대영역에 대한 직접 추정값이며, $\hat{Y}_{i..}$ 는 i 소지역을 * 추정법으로 추정한 것이다.

4.5 모형기반 간접추정법

4.5.1 기본적인 지역 수준 모형

소지역 추정시 모형에 근거한 추정방법이 많은 사람들의 관심을 끌고 있는 것은 다음과 같은 몇가지 장점에 기인한다. 먼저 모형 기반 추정법은 소지역들을 연결하고 있는 모형 구조가 소지역 간의 복잡한 오차구조를 내포하고 있기 때문에 소지역 간의 변동을 반영하여 소지역 추정의 정확도를 높일 수 있다는 점이며, 또한 표본자료로부터 모형의 유용성이 확인될 수 있고, 연속형의 자료뿐만 아니라 범주형 자료 및 시계열 자료와 같은 다양한 경우들에 대해서도 모형화하여 추론할 수 있으며, 모형 기반 추정법으로 소지역 추정량들과 연관있는 많은 측도들이 얻어질 수 있다는 장점들을 들 수 있다.

지역 간의 공변량을 포함하고 있는 지역 수준 모형을 이용하여 경험적 최량선형불편예측(EBLUP) 추정량, 경험적 베이스(EB) 추정량, 계층적 베이스(HB) 추정량에 대해 소개하기로 한다. 지역 수준 모형은 기본적으로 두가지의 성분들로 이루어진다. 즉, 소지역에 대한 직접추정량 $\hat{\theta}_i$ 과 소지역의 보조변수들로 표현되는 θ_i 의 두 가지 성분들을 모형으로 연결하여 모형 기반 추정량을 찾아내게 된다.

지정된 함수 $g(\cdot)$ 에 대하여 직접 추정량 $\hat{\theta}_i = g(\hat{Y}_i)$ 은 모집단의 값 $\theta_i = g(\bar{Y}_i)$ 와 표본추출오차 e_i 에 의해 다음과 같이 표현될 수 있다.

$$\hat{\theta}_i = \theta_i + e_i, \quad i=1, 2, \dots, m \quad (4.16)$$

여기에서 표본추출오차 e_i 는 서로 독립이며, 평균이 0, 분산이 ψ_i 임이 가정되며, 보통 ψ_i 는 기지인 값으로 가정된다.

θ_i 는 소지역의 정보를 나타내는 보조변수 $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})$ 를 이용하여 선형회귀모형을 통해 표현한다.

$$\begin{aligned}\theta_i &= z_{i1}\beta_1 + z_{i2}\beta_2 + \cdots + z_{ip}\beta_p + v_i \\ &= z_i^T \beta + v_i\end{aligned}\quad (4.17)$$

여기에서 모형오차 v_i 는 서로 독립이며, 평균이 0, 분산 σ_v^2 을 갖고, 표본 추출오차 e_i 와는 서로 독립임을 가정한다.

마지막으로 (4.16)과 (4.17)의 두 성분들을 결합하면 다음과 같은 결합 모형을 얻을 수 있다.

$$\hat{\theta}_i = z_i^T \beta + v_i + e_i \quad (4.18)$$

위의 결합모형은 고정효과 β 와 소지역 랜덤효과 v_i 를 갖는 선형혼합효과모형의 일종이며, 특히 설계 기반 확률변수(design-based random variable) e_i 와 모형 기반 확률변수(model-based random variable) v_i 를 동시에 포함하고 있는 모형이다. 여기에서 모수 σ_v^2 은 소지역들의 동질성을 나타내는 측도이다.

4.5.2 경험적 베이스(EB)방법

경험적 최량선형불편예측(EBLUP) 방법, 경험적 베이스(EB) 방법 및 계층적 베이스(HB) 방법은 모형에 근거한 소지역 추정문제에 많이 활용되고 있는 방법이다. 특히 경험적 최량선형불편예측 방법은 선형혼합모형을 이용한 추론에 응용되어 왔고, 경험적 베이스 방법 및 계층적 베이스 방법은 좀 더 일반적인 모형을 이용한 소지역 추정에 활용되고 있다.

EBLUP 추정량은 랜덤오차 e_i 와 v_i 의 분포에 대한 가정을 필요로 하지 않으나, MSE 추정을 위해 정규분포를 가정하기도 한다. 또한, EBLUP 추정량과 EB 추정량은 e_i 와 v_i 를 정규분포로 가정했을 경우에는 동일하며, HB 추정량과는 근사적으로 같게 나타난다. 그러나 추정량들의 변동을 나타내는 측도들은 동일하지는 않다.

고정계수 l_i 를 갖는 θ_i 의 선형추정량 $\sum_i l_i \hat{\theta}_i$ 가 모형 (4.18)에 대해서

$\sum_i k_i \hat{\theta}_i - \theta_i$ 의 기대값이 0을 만족할 때, $\sum_i k_i \hat{\theta}_i$ 를 θ_i 의 선형불편예측(LUP)추정량이라 한다. θ_i 의 최량선형불편예측(BLUP)추정량은 선형불편예측(LUP) 추정량들 중 최소평균제곱오차를 갖는 추정량을 말한다.

모형 (4.18) 하에서 θ_i 의 BLUP추정량은 다음과 같이 주어진다(Prasad and Rao,1990).

$$\tilde{\theta}_i(\sigma_v^2) = \gamma_i \hat{\theta}_i + (1 - \gamma_i) z_i^T \tilde{\beta}(\sigma_v^2) \quad (4.19)$$

여기에서 $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \psi_i)$ 이고, $\tilde{\beta}(\sigma_v^2)$ 은 가중치 $(\sigma_v^2 + \psi_i)^{-1}$ 을 갖는 가중최소제곱추정량으로 아래와 같이 주어진다.

$$\tilde{\beta}(\sigma_v^2) = \left(\sum_i \gamma_i z_i z_i^T \right)^{-1} \left(\sum_i \gamma_i z_i y_i \right) \quad (4.20)$$

(4.19)식의 BLUP추정량은 가중치 γ_i 를 갖는 직접추정량 $\hat{\theta}_i$ 과 가중치 $1 - \gamma_i$ 를 갖는 회귀합성추정량 $z_i^T \tilde{\beta}(\sigma_v^2)$ 의 가중결합으로 볼 수 있다. 또한, 표본분산 ψ_i 가 작을때(σ_v^2 이 클 경우) BLUP추정량은 직접추정량 $\hat{\theta}_i$ 에 큰 가중치가 부여되고, 반대의 경우에는 회귀합성추정량 $z_i^T \tilde{\beta}(\sigma_v^2)$ 에 큰 가중치가 부여된다. 표본이 추출되지 않은 지역들에 대해서는 BLUP추정량은 회귀합성추정량만으로 주어질 수 있다.

BLUP추정량의 변동의 측도는 추정량의 $MSE = E(est. - \theta_i)^2$ 에 의해 주어지며 다음과 같다.

$$MSE(\tilde{\theta}_i(\sigma_v^2)) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) , \quad (4.21)$$

여기에서

$$g_{1i}(\sigma_v^2) = \gamma_i \psi_i ,$$

$$g_{2i}(\sigma_v^2) = \sigma_v^2 (1 - \gamma_i)^2 z_i^T \left(\sum_i \gamma_i z_i z_i^T \right)^{-1} z_i$$

로 주어진다.

식(4.19)와 (4.21)은 랜덤오차 v_i 와 e_i 에 관한 분포의 가정을 필요로 하지는 않는다.

주요 항 $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$ 는 $O(1)$, $g_{2i}(\sigma_v^2)$ 은 $O(m^{-1})$ 에 유계인 항이며, 이로부터 $BLUP$ 추정량의 MSE 값은 γ_i 나 모형분산 σ_v^2 이 표본분산 ψ_i 에 비해 작을 경우 직접추정량의 MSE 값보다 훨씬 작아질 수 있다는 사실을 알 수 있다. 따라서 소지역 추정의 정확도는 표본 분산에 비해 모형 분산을 작게 할 수 있는 보조변수에 크게 의존한다고 볼 수 있다.

대부분의 문제에서는 모형분산 σ_v^2 은 미지이므로 적절한 $\hat{\sigma}_v^2$ 을 추정하여 $EBLUP$ 추정량 $\hat{\theta}_i = \hat{\theta}_i(\hat{\sigma}_v^2)$ 을 산출한다. 이때 소지역의 평균 \bar{Y}_i 의 추정량은 $g^{-1}(\hat{\theta}_i)$ 로, σ_v^2 의 추정량은 $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$ 로 주어진다. 여기에서 $\tilde{\sigma}_v^2$ 은 다음 식을 만족한다.

$$(m-p)\tilde{\sigma}_v^2 = \sum_i (\hat{\theta}_i - z_i^T \beta^*)^2 - \sum_i \psi_i h_{ii} \quad (4.22)$$

(4.22)식에서 $h_{ii} = z_i^T (\sum_i z_i z_i^T)^{-1} z_i$ 이고, β^* 는 β 의 OLS (ordinary least squares) 추정량이다. 한편, $\tilde{\sigma}_v^2$ 은 다음과 같은 비선형 방정식의 반복적인 해로써 구할 수도 있다.

$$a(\sigma_v^2) = \sum_i \{ \hat{\theta}_i - z_i^T \tilde{\beta}(\sigma_v^2) \}^2 / (\sigma_v^2 + \psi_i) = m-p \quad (4.23)$$

여기에서 $\tilde{\beta}(\sigma_v^2)$ 은 (4.20)식에 주어졌고, (4.23)식의 가운데 항은 가중잔차 제곱합, $m-p$ 는 가중잔차제곱합과 관계가 되는 자유도이다. 만약 $\hat{\sigma}_v^2 = 0$ 이면 $EBLUP$ 추정량 $\hat{\theta}_i$ 는 회귀합성추정량 $z_i^T \hat{\beta}$ 로 축약된다. 단, $\hat{\beta} = \tilde{\beta}(\hat{\sigma}_v^2)$ 이며, (4.20)식에서 σ_v^2 대신에 $\hat{\sigma}_v^2$ 으로 대체하여 산출한다. 물론 위의 (4.22), (4.23)식으로부터 얻게 되는 추정량들도 v_i 와 e_i 의 분포에

대한 가정을 필요로 하지는 않는다.

만약 랜덤오차 v_i 와 e_i 가 정규분포를 따른다고 가정한다면, $\hat{\theta}_i$ 는 평균이 $z_i^T \beta$ 이고 분산이 $\sigma_v^2 + \psi_i$ 인 서로 독립인 정규분포를 따르게 된다. 이러한 분포에 대한 가정 하에서 계산된 β 와 σ_v^2 의 최대우도추정량을 제한 최대우도추정량(REML)이라 하며, 선형혼합모형 하에서도 근사적으로 유효하다. 따라서 $\hat{\theta}_i$ 의 BLUP추정량을 산출 시 σ_v^2 의 REML추정량을 이용하여도 근사적으로 타당하다.

경험적 베이지(EB) 추정법은 랜덤오차 v_i 와 e_i 가 정규분포를 따른다는 가정 하에서 출발한다. $(\hat{\theta}_i, \theta_i)$ 의 결합분포가 평균이 $(z_i^T \beta, z_i \beta)$, 분산이 $(\sigma_v^2 + \psi_i, \sigma_v^2)$, 상관계수가 γ_i 인 이변량 정규분포를 따른다고 가정하자. 이때, θ_i 의 평균제곱오차를 최소로 하는 베이지 추정량은 다음과 같다.

$$\begin{aligned} \hat{\theta}_i^B(\beta, \sigma_v^2) &= E(\theta_i | \hat{\theta}_i, \beta, \sigma_v^2) \\ &= \gamma_i \hat{\theta}_i + (1 - \gamma_i) z_i^T \beta \end{aligned} \quad (4.24)$$

(4.24)식의 베이지 추정량은 선형성 또는 불편성을 만족하지는 않는다. 여기에서 모수 β 와 σ_v^2 을 제한최대우도(REML)추정량으로 대체하여 다음과 같은 θ_i 의 경험적 베이지(EB) 추정량을 얻는다.

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B(\hat{\beta}, \hat{\sigma}_v^2) \quad (4.25)$$

경험적 베이지(EB) 추정량 $\hat{\theta}_i^{EB}$ 는 정규분포의 가정 하에서는 EBLUP 추정량 $\hat{\theta}_i$ 와 같다. 그러나 경험적 베이지 방법은 $\hat{\theta}_i$ 과 θ_i 의 임의의 결합 분포에 대해서도 일반적으로 응용할 수 있다는 점을 장점으로 들 수 있다.

EBLUP추정량 $\hat{\theta}_i = \hat{\theta}_i(\hat{\sigma}_v^2)$ 의 MSE추정량은 (4.21)식에서 σ_v^2 대신 $\hat{\sigma}_v^2$ 을 대체하여 얻어질 수 있으나, 이 경우에는 σ_v^2 에 대한 추정효과가 무시

되기 때문에 MSE 의 추정값은 과소추정되는 경향을 보인다. 이러한 문제 때문에 Prasad and Rao(1990)는 $\{u_i\}$ 와 $\{e_i\}$ 에 대해 정규성을 가정하여 근사적으로 불편인 $EBLUP$ 추정량 $\tilde{\theta}_i$ 의 MSE 추정량을 제안하였다. Prasad and Rao(1990)가 제안한 MSE 추정량은 (4.22)식의 σ_v^2 의 적률추정량을 사용하였을 경우 다음과 같이 주어진다.

$$mse(\tilde{\theta}_i) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) \quad (4.26)$$

여기에서

$$g_{1i}(\sigma_v^2) = \gamma_i \psi_i,$$

$$g_{2i}(\sigma_v^2) = \sigma_v^2 (1 - \gamma_i)^2 z_i^T \left(\sum_j \gamma_j z_j z_j^T \right)^{-1} z_i,$$

$$g_{3i}(\sigma_v^2) = [\psi_i^2 / (\sigma_v^2 + \psi_i)^3] h(\sigma_v^2),$$

$$h(\sigma_v^2) = 2m^{-2} \sum_j (\sigma_v^2 + \psi_j)^2$$

로 주어진다.

최근들어 Jing, Lahiri and Wan(1999)은 근사적으로 불편인 잭나이프 MSE 추정량을 제안하였다. 잭나이프 방법은 랜덤인 지역효과들을 갖는 로지스틱 회귀와 같은 좀 더 복잡한 모형들에 대해서도 쉽게 적용할 수 있다는 장점을 갖고 있다.

θ_i 의 EB 추정량 (4.25)를 $\tilde{\theta}_i^{EB} = k(\hat{\theta}_i, \hat{\phi})$ 로 표현할 때, 잭나이프 절차는 다음과 같다. 여기에서 $\phi = (\beta, \sigma_v^2)$ 은 모형에서의 모수 β 와 σ_v^2 을 나타낸다.

(i) l 번째 지역의 자료 $(\hat{\theta}_l, z_l)$ 을 제외한 ϕ 의 추정량 $\hat{\phi}(l)$ 을 계산한다. 이때의 EB 추정량을 $\tilde{\theta}_i^{EB}(l) = k(\hat{\theta}_i, \hat{\phi}(l))$ 로 나타내자.

(ii) $\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^m [\tilde{\theta}_i^{EB}(l) - \tilde{\theta}_i^{EB}]^2$ 를 계산한다.

(iii) $\hat{M}_{1i} = g_{1i}(\hat{\sigma}_v^2) - \frac{m-1}{m} \sum_{l=1}^m [g_{1i}(\hat{\sigma}_v^2(l)) - g_{1i}(\hat{\sigma}_v^2)]^2$ 을 계산한다.

(iv) 잭나이프 MSE 추정량 $mse_j(\hat{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$ 를 계산한다.

\hat{M}_{1i} 은 ϕ 가 기지일 때 MSE 에 대한 추정량이며, \hat{M}_{2i} 는 모형 모수 ϕ 를 추정할 때 추가적으로 발생하는 MSE 에 대한 변화량을 추정한다.

4.5.3 계층적 베이즈(HB) 방법

계층적 베이즈(HB) 방법을 이용한 추론은 비교적 추론의 정확도가 높고, 복잡한 유형의 문제들에서도 최근에 개발된 $MCMC$ (Monte Carlo Markov Chain)방법을 이용하여 해결할 수 있다. 깃스 샘플러가 이러한 문제 해결을 위해 제공된다. HB방법에서는 모형 모수 $\phi = (\beta, \sigma_v^2)$ 뿐만 아니라 모집단의 값 θ_i 가 랜덤으로 간주되며, 모형 모수들에 대한 사전분포가 명시된다. θ_i 의 추론은 주변 사후분포에 의해 결정된다. 즉, 주어진 자료 $\{(\hat{\theta}_i, z_i), i=1, 2, \dots, m\}$ 에 대한 조건부 분포 $f(\theta_i | \hat{\theta})$ 에 의해 추론이 행해진다. 여기에서 $\hat{\theta}$ 은 직접추정값 $\hat{\theta}_i$ 의 벡터이다. 특히 θ_i 는 사후분포의 평균 $E(\theta_i | \hat{\theta})$ 에 의해 추정되며, 추정량의 변동은 사후분포의 분산 $V(\theta_i | \hat{\theta})$ 에 의해 추정된다.

먼저 σ_v^2 이 기지인 상태를 가정하고 β 에 관한 사전분포를 배정하기로 한다. β 의 사전분포가 상수에 비례하고(i.e improper prior), v_i 와 e_i 가 정규분포를 따른다고 가정한다면, 이때 사후평균 $E(\theta_i | \hat{\theta}, \sigma_v^2)$ 은 (4.19)식의 $BLUP$ 추정량 $\tilde{\theta}_i(\sigma_v^2)$ 과 동일하다. 더욱이 사후분산 $V(\theta_i | \hat{\theta}, \sigma_v^2)$ 은 (4.21)식의 $BLUP$ 추정량의 MSE 와 같다. 따라서 σ_v^2 이 기지인 상태에서는 HB 방법과 $EBLUP$ 방법은 동일한 추론을 이끌어 낸다고 볼 수 있다.

실제의 문제에서는 σ_v^2 은 대부분 미지의 값으로 나타난다. 이러한 경우

에는 β 뿐만 아니라 σ_v^2 에 관한 사전분포를 고려해야 하며, 또한 서로 독립임을 가정하여 주변사후분포 $f(\sigma_v^2 | \hat{\theta})$ 을 이끌어 낸다. 만약 σ_v^2 에 관한 사전분포를 불완전(improper) 사전분포로 배정한다면, θ_i 의 사후분포가 불완전 사후분포가 될 수 있기 때문에 이러한 문제를 피하기 위해서 $\tau_v = \sigma_v^{-2}$ 의 사전분포를 $G(a, b)$ 와 같이 배정 한다 (여기서 $G(a, b) : f(\tau_v) \propto \exp(-a\tau_v) \tau_v^{b-1}$). 주변사후분포 $f(\sigma_v^2 | \hat{\theta})$ 를 이용한 HB 추정량 $E(\theta_i | \hat{\theta})$ 은 다음 식과 같이 주어진다.

$$\hat{\theta}_i^{HB} = E(\theta_i | \hat{\theta}) = \int \tilde{\theta}_i(\sigma_v^2) f(\sigma_v^2 | \hat{\theta}) d\sigma_v^2 \quad (4.27)$$

위의 (4.27)식을 $E_{\sigma_v^2 | \hat{\theta}}[\tilde{\theta}_i(\sigma_v^2)]$ 으로 표현하면, 사후분산 $V(\theta_i | \hat{\theta})$ 은 다음과 같다.

$$V(\theta_i | \hat{\theta}) = E_{\sigma_v^2 | \hat{\theta}}[g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2)] + V_{\sigma_v^2 | \hat{\theta}}[\tilde{\theta}_i(\sigma_v^2)] \quad (4.28)$$

여기에서 $V_{\sigma_v^2 | \hat{\theta}}$ 는 $f(\sigma_v^2 | \hat{\theta})$ 에 관한 분산을 의미한다.

위에서 소개한 (4.27)식과 (4.28)식은 일차원 수치적분으로 계산된다. 좀 더 복잡한 모형에 대한 고차원 수치적분은 MCMC방법을 이용하여 계산할 수 있다. (4.27)식으로부터 $\hat{\theta}_i^H$ 는 EBLUP(EB)추정량 $\tilde{\theta}_i(\hat{\sigma}_v^2)$ 과 근사적으로 같다는 것을 알 수 있다.

깁스 샘플링은 위의 (4.27)식와 (4.28)식을 결정하기 위해 사용될 수 있는 일종의 MCMC방법이다. 깁스 샘플링을 수행하기 위해서는 다음과 같은 깁스 조건부 분포들이 필요하다.

$$(i) [\beta | \theta, \sigma_v^2, \hat{\theta}] \sim N_p \left[\left(\sum_i z_i z_i^T \right)^{-1} \left(\sum_i z_i \theta_i \right), \sigma_v^2 \left(\sum_i z_i z_i^T \right)^{-1} \right]$$

$$(ii) [\theta_i | \beta, \sigma_v^2, \hat{\theta}] \sim N[\tilde{\theta}_i^B(\beta, \sigma_v^2), g_{1i}(\sigma_v^2) = \gamma_i \psi_i]$$

$$(iii) [\tau_v = \sigma_v^{-2} | \beta, \theta, \hat{\theta}] \sim G(\bar{a}, \bar{b}),$$

$$\text{단, } \bar{a} = \frac{1}{2} \sum_i (\theta_i - z_i^T)^2 + a, \quad \bar{b} = \frac{m}{2} + b.$$

깁스 알고리즘은 다음 절차에 의해 이루어진다.

(a) $\theta_i = \theta_i^{(0)}, \sigma_v^2 = \sigma_v^{2(0)}$ 을 초기값으로 하여 (i)로부터 $\beta^{(1)}$ 을 계산

(b) $\beta = \beta^{(1)}, \sigma_v^2 = \sigma_v^{2(0)}$ 을 이용하여 (ii)로부터 $\theta_i^{(1)}$ 을 계산

$$(i = 1, 2, \dots, m)$$

(c) $\theta_i = \theta_i^{(1)}$ 과 $\beta = \beta^{(1)}$ 을 이용하여 (iii)으로부터 $\sigma_v^{2(1)}$ 을 계산

(d) 절차 (a), (b), (c)를 한 사이클로 하여 반복 수행

수렴이 이루어지는 시점 t 까지 충분히 반복한 후, 이 후부터 얻어지는 J 개의 표본 $\{\beta^{(t+j)}, \sigma_v^{2(t+j)}, \theta_1^{(t+j)}, \dots, \theta_m^{(t+j)}; j = 1, 2, \dots, J\}$ 을 $\beta, \sigma_v^2, \theta_1, \dots, \theta$ 의 결합 사후분포로 얻은 표본으로 간주한다. 초기값은 보통 $\theta_i^{(0)} = \bar{\theta}_i^E, \sigma_v^{2(0)} = \sigma_v^2$ 의 REML 추정량을 사용한다.

위에서 계산된 J 개의 표본을 이용하여 θ_i 의 사후평균, 사후분산을 다음과 같이 추정한다.

$$\begin{aligned} \bar{\theta}_i^{HB} &\approx \frac{1}{J} \sum_j \bar{\theta}_i(\sigma_v^{2(t+j)}) \\ &= \frac{1}{J} \sum_j \bar{\theta}_i(j) \\ &= \bar{\theta}_i(\cdot), \end{aligned} \tag{4.29}$$

$$V(\theta_i | \hat{\theta}) \approx \frac{1}{J} \sum_j [g_{1i}(\sigma_v^{2(t+j)}) + g_{2i}(\sigma_v^{2(t+j)})] + \frac{1}{J} \sum_j [\bar{\theta}_i(j) - \bar{\theta}(\cdot)]^2 \tag{4.30}$$

제5장 시군구 경황통계 작성

5.1 개 요

통계청에서는 취업, 실업 등과 같은 경제적 특성을 조사하여 국가의 고용정책 입안과 평가에 필요한 기초자료를 수집할 목적으로 매월 3만 표본 가구 내에 거주하는 만 15세 이상인 사람들을 대상으로 경제활동인구 조사를 실시하고 있다. 매월 15일을 포함하는 주 중에 표본 가구 내에 거주하는 사람들의 취업, 실업 및 비 경제활동인구 관련 사항을 방문면접이나 컴퓨터면접 방식으로 조사한 후 조사된 자료를 직접 추정방법에 의해 추정 계산하여 매 익월 말경에 대영역인 7개 광역시와 9개 도에 대해 조사 결과를 발표하고 있다.

1995년부터 시작된 지방자치제도와 1997년 IMF 사태에 기인하여 최근에는 실업 관련 통계뿐만 아니라 다양한 분야의 통계 작성도 시군구의 소지역 단위까지 작성해야 한다는 인식이 높아지고는 있지만 현재와 같은 대영역 표본설계를 기반으로 하는 통계 작성 방법으로는 신뢰성 있는 소지역 통계 생산은 불가능한 실정이다. 왜냐하면 소지역 통계의 작성은 단순히 추정단계에서 추정량의 선택을 통해서 해결될 수 있는 것이 아니라 통계조사의 계획, 표본설계, 추정 등 통계조사 전 과정을 종합적으로 고려할 때 가능한 일이기 때문이다.

통계청의 경제활동인구조사는 대영역인 광역시와 도별 통계를 목적으로 표본설계 되었기 때문에 소지역인 시군구 단위는 표본설계에 반영된 관심영역이 아니다. 따라서 현재 활용되고 있는 대영역 기반의 표본설계를 이용하여 소지역 통계를 직접 생산할 경우 시군구 지역에 배정된 표본조사구 수가 불균형적이고 특정 시군구 단위의 소지역에 대해서는 표본조사구의 수가 하나 내지 두 개 정도로 너무 작게 배정되어 있기 때문에 신뢰할만한 소지역 단위의 통계생산은 어렵게 된다.

우리의 목적은 대영역 기반 표본설계의 구조 하에서 직접 생산된 시군구 단위의 직접추정값들을 소지역 추정에 이용되는 명시적인 모형을 통해

보정하여 어느 정도 신뢰성을 확보하는 일이다. 이를 위하여 사전에 소지역에 대한 층화, 표본 배정, 집락화의 수준에 대한 검토 후 직접추정값을 보정할 수 있는 다양한 보조정보들을 고려하였다.

이 보고서에서는 현재 통계청에서 실시하고 있는 대영역 기반 경제활동인구조사의 표본설계 하에서 직접적으로 생산된 시군구 단위의 소지역에 대한 직접추정값들을 합성추정법 또는 복합추정법과 같은 설계 기반 소지역 추정법을 통해 보정하여 시군구의 실업자 총계를 추론한다. 표본조사구 수가 불균형으로 배정된 상태에서 추정된 소지역의 월별 직접추정값들에 대해 센서스 및 행정보고자료를 통해 선택된 보조정보를 이용하여 직접추정값들에 대한 보정을 시도하였다.

“Borrow Strength”를 적용하기 위하여 대영역별 경제활동인구조사 자료를 크게 시군구 그룹으로 크게 3개 그룹으로 분할하고, 각 그룹 내에서는 유사성질 범주를 성별(남, 여)-연령대별(15-29세, 30세 이상)로 총 4개의 범주로 구분하였다. 각 그룹별 4개 범주에 대해서는 범주별 실업자 총계를 산출하여 대영역 내의 시군구 단위에 대한 실업자 총계 추정을 위한 보조정보로 활용하였다.

5.2 소지역 추정을 위한 집락화 방안

경제활동인구조사는 대영역인 특·광역시 및 도지역의 총 16개 대영역에 대한 경제활동인구 특성을 파악할 목적으로 표본설계 되었다. 현재의 경제활동인구조사에서 특·광역시 및 도지역에 배정된 표본조사구 수는 총 1,629개이며 이들 대영역에 대한 표본조사구 수의 분포는 다음 <표 5.1>과 같다.

<표5.1> 특·광역시 및 도지역의 표본조사구 수 분포

시도	주민등록 인구	시군구 수	표본조사구		시도	주민등록 인구	시군구 수	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
서울	10,331,244	25	207	12.71	충북	1,504,518	12	81	4.97
부산	3,786,033	16	117	7.18	충남	1,928,088	16	90	5.52
대구	2,539,587	8	90	5.52	전북	2,013,923	14	81	4.97
인천	2,581,557	10	108	6.63	전남	2,104,052	22	90	5.52
광주	1,387,360	5	81	4.97	경북	2,802,597	23	108	6.63
대전	1,408,809	5	81	4.97	경남	3,124,123	20	99	6.08
울산	1,060,378	5	63	3.87	제주	547,964	4	45	2.76
경기	9,612,036	31	198	12.15	합계	48,289,173	234	1,629	100.00
강원	1,556,904	18	90	5.52					

* 주민등록인구는 2001년 12월 기준

<표5.1>에서 대영역 내의 시군구 단위 소지역들의 수는 총 234개이며 이 들 시군구 단위 소지역들 각각에 대해 실업자 총계를 추정할 경우 시군구 단위 소지역들에 배정된 표본조사구의 수가 충분하다면 이러한 소지역들의 실업자 총계 추정값들은 신뢰성을 확보할 수 있다. 그러나 경제활동인구조사의 표본설계는 대영역의 경제활동 특성을 파악할 목적으로 설계되었기 때문에 시군구 단위의 소지역들에 배정된 표본조사구의 수는 매우 불균형적인 분포를 나타낸다. 따라서 표본조사구 수가 충분히 배정되지 못한 특정 소지역들의 경우 실업자 총계 추정값들의 추정오차는 신뢰할 수 없을 정도로 매우 큰 값을 나타낼 가능성이 있다. 표본조사구 수가 충분하지 못한 시군구단위 소지역 추정값들의 정확성을 제고하기 위한 방안으로 대영역 내의 시군구단위 소지역들을 연구변수들에 의해 집락화하여 특성이 유사한 시군구단위 소지역들을 판별한 후 이들을 합성하여 실업자 총계를 추정하는 방안이 고려될 수 있다.

대영역인 강원도의 시군단위 소지역들을 예를 들어 보자. 다음 <표 5.2>는 강원도 내의 18개 시군단위 소지역들에 대한 표본조사구 수의 분포를 나타낸다.

<표5.2> 강원도 시군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	춘천시	252,019	13	14.44	11	고성군	34,728	3	3.33
2	홍천군	74,224	5	5.56	12	양양군	30,276	3	3.33
3	철원군	52,224	3	3.33	13	동해시	103,933	5	5.56
4	화천군	25,026	3	3.33	14	태백시	56,254	3	3.33
5	양구군	23,274	2	2.22	15	삼척시	79,862	4	4.44
6	인제군	33,133	3	3.33	16	영월군	52,151	3	3.33
7	원주시	275,217	15	16.67	17	평창군	46,564	3	3.33
8	횡성군	45,701	2	2.22	18	정선군	49,111	3	3.33
9	강릉시	232,664	13	14.44	합계		1,556,904	90	100.00
10	속초시	90,543	4	4.44					

표본조사구 수가 4개 이상 배정된 시군단위 소지역들에 대한 실업자 총계 추정값들은 앞으로 소개될 소지역추정법인 5.3절의 합성추정법과 5.4절의 복합추정법을 적용하면 추정값들의 신뢰성은 상당 부분 확보될 수 있다. 그러나 표본조사구 수가 3개 이하인 철원군, 화천군, 양구군, 인제군, 횡성군, 고성군, 양양군, 태백시, 영월군, 평창군, 정선군의 11개 시군단위 소지역들은 5.3절과 5.4절에서 소개되는 소지역추정법(합성추정법, 복합추정법)을 적용하더라도 실업자 총계 추정값의 추정오차는 매우 큰 값을 나타낸다.

이와 같이 대영역 내에서 표본조사구 수가 충분하지 못한 시군단위 소지역들의 실업자 총계를 추정할 때 다음과 같은 대안이 고려될 수 있다. 예를 들어 표본조사구 수가 3개 이하인 횡성군(2개), 영월군(3개)과 고성군(3개)의 경우, 이들 세지역의 실업률이 매우 유사하다고 가정할 수 있다면, 이 때 횡성군, 영월군과 고성군을 하나의 그룹으로 묶어 표본조사구 수를 충분히 확보한 다음 실업자 총계를 좀 더 안정적으로 추정한 후 각 지역별 상주추정인구비례로 횡성군, 영월군과 고성군에 각각 추정값을 배분하여 주는 방법을 생각할 수 있다. 이러한 방법은 표본조사구 수가 적

은 소지역들에 대한 실업자 총계 추정 시 매우 합리적인 대안으로 이용될 수 있다. 이 때 대영역 내에서 실업률이 유사한 시군단위 소지역들을 그룹으로 묶는 방법은 집락분석 결과를 이용할 수 있다.

대영역 내의 시군구단위 소지역들을 실업률이 유사한 동질적인 그룹들로 집락화하는 방법을 소개하면 다음과 같다. 집락분석에 이용된 자료는 보다 정확한 분석을 위해 2000년 「인구주택 총 조사」 자료를 이용하였다. 「인구주택 총 조사」 자료에서 시군구단위 소지역들의 “성별(남, 여)-연령대별(15-29세, 30세 이상)”의 4개 범주의 실업률 X_1, X_2, X_3, X_4 를 계산하여 대영역 내의 시군구단위 소지역들을 집락화하기 위한 연구변수들로 이용하였다. 대영역 A 내의 시군구 단위 소지역들의 수를 N_A 라 할 때, 데이터 셀(“templ.txt”)의 구조는 다음과 같다.

<u>“templ.txt”</u>	AREA	X_1	X_2	X_3	X_4
	1	x_{11}	x_{12}	x_{13}	x_{14}
	2	x_{21}	x_{22}	x_{23}	x_{24}
	⋮				
	N_A	$x_{N_A,1}$	$x_{N_A,2}$	$x_{N_A,3}$	$x_{N_A,4}$

여기에서 x_{ij} 는 i 시군구 소지역에서 j 범주의 실업률을 나타낸다.

SAS의 FASTCLUS 프로시저를 이용하여 집락분석 프로그램을 작성하면 다음과 같다.

집락화 프로그램

```

DATA temp;
  INFILE 'A:\templ.txt';
  INPUT AREA X1-X4;
PROC FASTCLUS DATA=temp OUT=result
  MAXCLUSTERS= $n$  MAXITER=10 CONV=0.02;
  VAR X1-X4;
PROC PRINT DATA=result;
RUN;

```

FASTCLUS 처리절차는 유클리드 거리를 기초로 반복적 알고리즘을 통해 상호 배반적 집락을 형성시키며 연구의 목적에 적합한 동질적인 시군구단위들의 개수(MAXCLUSTERS= n)로 집락화 하는데 상당히 효과적인 방법으로 알려져 있다.

강원도 지역에 대해 집락의 개수를 MAXCLUSTERS=4로 지정하여 FASTCLUS 프로시저를 시행하면 다음 <표5.3>과 같은 결과를 얻을 수 있다.

<표5.3> 강원도 시군단위 소지역들의 집락분석 결과
():표본조사구 수

구 분	집락1	집락2	집락3	집락4
시군구	철원군(3), 화천군(3), 양구군(2), 인제군(3)	강릉시(13), 삼척시(4), 횡성군(2), 영월군(3), 고성군(3)	원주시(15), 동해시(5), 태백시(3), 속초시(4)	춘천시(13), 홍천군(5), 평창군(3), 정선군(3), 양양군(3)

집락분석 결과에 의하면 강원도의 시군단위 소지역들은 실업률이 유사한 4개의 집락들로 구분되었다. 따라서 위의 집락분석 결과를 참조하여 표본조사구 수가 3개 이하인 “철원군(3개)”, “화천군(3개)”, “양구군(2개)”, “인제군(3개)”, “횡성군(2개)”, “영월군(3개)”, “고성군(3개)”, “태백시(3개)”, “평창군(3개)”, “정선군(3개)”과 “양양군(3개)”의 11개 시군단위 소지역들을 실업률이 유사한 소지역들끼리 묶어 그룹별 실업자 총계를 추정한 후 소지역별 상주추정인구비례로 추정값을 배분한다면 좀 더 정확한 추정결과를 얻을 수 있다. <표5.3>의 집락1에서 “철원군”, “화천군”, “양구군”과 “인제군”의 소지역들은 실업률이 유사한 동일집락 내에 속한다. 표본조사구 수와 집락분석 결과를 참조하여 이들 4개 소지역들을 “철원군(3)+화천군(3)”과 “양구군(2)+인제군(3)”의 그룹으로 묶어 충분한 표본조사구 수를 확보한 후 5.3절과 5.4절에서 소개된 소지역추정법을 적용한다면 보다 안정적으로 실업자 총계를 추정할 수 있고 추정오차를 줄일 수 있다. 같은

방법으로 “횡성군(2)+영월군(3)+고성군(3)”을 하나의 그룹으로, “평창군(3)+정선군(3)+양양군(3)”을 또 다른 하나의 그룹으로 묶어 실업자 총계를 추정할 수 있다. 시지역의 월별 실업률은 대체적으로 안정적인 추정결과를 나타내므로 “태백시(3개)”는 다른 유사 지역과 묶지 않고 단독으로 추정한다.

시군구단위 소지역들에 배정된 표본조사구 수와 대영역 내의 소지역들에 대한 집락분석결과를 참조하여 표본조사구 수가 3개 이하인 시군구단위 소지역들을 경찰 특성이 유사한 그룹으로 합성하면 다음과 같이 요약할 수 있다.

① 서울특별시

<표5.4> 서울특별시 구단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록인구	표본조사구		순번	시군구	주민등록인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	종로구	187,038	4	1.93	14	마포구	382,195	8	3.86
2	중구	146,335	4	1.93	15	양천구	486,095	10	4.83
3	용산구	250,550	4	1.93	16	강서구	523,542	10	4.83
4	성동구	343,471	7	3.38	17	구로구	417,453	8	3.86
5	광진구	390,090	8	3.86	18	금천구	263,061	6	2.90
6	동대문구	383,822	8	3.86	19	영등포구	409,920	8	3.86
7	중랑구	449,965	9	4.35	20	동작구	407,793	8	3.86
8	성북구	453,517	10	4.83	21	관악구	529,741	11	5.31
9	강북구	352,317	7	3.38	22	서초구	397,983	7	3.38
10	도봉구	372,318	7	3.38	23	강남구	546,038	12	5.80
11	노원구	648,615	12	5.80	24	송파구	658,242	13	6.28
12	은평구	469,242	9	4.35	25	강동구	490,585	10	4.83
13	서대문구	371,316	7	3.38		합계	10,331,244	207	100.00

서울특별시 내의 모든 구단위 소지역들은 자체적으로 4개 이상의 표본조사구 수를 확보하고 있으므로 유사 소지역 합성은 고려하지 않고 소지역별로 독립적으로 실업자 총계를 추정한다. 단, 소지역추정법을 적용할 경우 발생할 수 있는 잠재적인 편향 가능성을 줄이기 위해 서울특별시를

“지리적”, “산업특성별” 또는 “인구분포 특성별”로 유사성질을 갖는 동질적인 4개의 그룹으로 분할하여 각각의 그룹별로 소지역추정법을 적용하여 실업자 총계 추정값들을 산정한다.

서울 A지구(강북동 지역)

성동구, 광진구, 중랑구, 성북구, 강북구, 도봉구, 노원구

서울 B지구(강북서 지역)

종로구, 중구, 용산구, 은평구, 서대문구, 마포구, 동대문구

서울 C지구(강남동 지역)

서초구, 강남구, 송파구, 강동구

서울 D지구(강남서 지역)

양천구, 강서구, 구로구, 금천구, 영등포구, 동작구, 관악구

② 부산광역시

<표5.5> 부산광역시 구/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	중구	57,658	3	2.56	10	사하구	385,967	11	9.40
2	서구	152,752	5	4.27	11	금정구	287,730	9	7.69
3	동구	125,085	4	3.42	12	강서구	60,773	3	2.56
4	영도구	180,626	6	5.13	13	연제구	226,768	7	5.98
5	부산진구	427,511	13	11.11	14	수영구	181,777	5	4.27
6	동래구	295,915	9	7.69	15	사상구	302,349	9	7.69
7	남구	303,301	9	7.69	16	기장군	76,721	3	2.56
8	북구	311,893	9	7.69	합계		3,786,033	117	100.00
9	해운대구	409,207	12	10.26					

부산광역시에서 표본조사구 수가 3개 이하인 지역은 “중구(3개)”, “강서구(3개)”와 “기장군(3개)”의 3개 지역이다.

<표5.6> 부산광역시 구/군단위 소지역들의 집락분석 결과

(): 표본조사구 수

구 분	집락1	집락2	집락3	집락4
시군구	중구(3), 동래구(9), 기장군(3)	서구(5), 영도구(6), 남구(9), 사하구(11), 금정구(9),연제구(7), 수영구(5)	동구(4), 부산진구(13), 북구(9), 해운대구(12), 사상구(9)	강서구(3)

같은 집락에 속하는 “중구(3개)”와 “기장군(3개)”을 “중구(3개)+기장군(3개)”의 그룹으로 묶어 추정한다. “강서구(3개)”의 실업률은 다른 구단위 소지역들과 매우 상이한 특성을 보이므로 이 지역은 독립적으로 실업자 총계를 추정한다.

부산광역시도 서울시와 마찬가지로 소지역추정값들의 편향과 추정오차를 줄이기 위해 경찰 특성이 유사한 2개의 동질적인 그룹으로 구분하여 소지역추정법을 적용한다.

부산 A지구

중구, 동래구, 기장군, 서구, 영도구, 남구, 사하구, 금정구, 연제구, 수영구

부산 B지구

동구, 부산진구, 북구, 해운대구, 사상구, 강서구

③ 대구광역시

<표5.7> 대구광역시 구/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	중구	90,097	4	4.44	6	수성구	457,704	15	16.67
2	동구	336,999	12	13.33	7	달서구	604,371	21	23.33
3	서구	282,847	11	12.22	8	달성군	158,278	5	5.56
4	남구	193,064	8	8.89	합계	2,539,587	90	100.00	
5	북구	416,227	14	15.56					

대구광역시는 모든 구 및 군단위 소지역들이 자체적으로 4개 이상의 표본조사구 수를 확보하고 있으므로 유사 소지역합성을 고려하지 않고 소지역별로 독립적으로 실업자 총계를 추정한다.

④ 인천광역시

<표5.8> 인천광역시 구/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	중구	75,278	4	3.70	7	계양구	340,830	14	12.96
2	동구	76,223	4	3.70	8	서구	351,598	15	13.89
3	남구	427,339	17	15.74	9	강화군	67,067	4	3.70
4	연수구	261,851	11	10.19	10	옹진군	14,068	0	0.00
5	남동구	413,342	17	15.74	합계	2,581,557	108	100.00	
6	부평구	553,961	22	20.37					

인천광역시는 “옹진군(0개)”을 제외한 모든 구 및 군단위 소지역들이 자체적으로 4개 이상의 표본조사구 수를 확보하고 있으므로 유사 소지역합성을 고려하지 않고 소지역별로 독립적으로 추정한다. “옹진군(0개)”은 배정된 표본조사구 수가 없으므로 분석에서 제외한다.

⑤ 광주광역시

<표5.9> 광주광역시 구단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	동구	119,556	7	8.64	4	북구	475,314	29	35.80
2	서구	301,302	17	20.99	5	광산구	264,734	14	17.28
3	남구	226,454	14	17.28	합계		1,387,360	81	100.00

광주광역시는 모든 구단위 소지역들이 자체적으로 4개 이상의 충분한 표본조사구 수를 확보하고 있으므로 유사 소지역합성을 고려하지 않고 소지역별로 독립적으로 추정한다.

⑥ 대전광역시

<표5.10> 대전광역시 구단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	동구	247,539	15	18.52	4	유성구	173,509	9	11.11
2	중구	269,300	16	19.75	5	대덕구	231,886	14	17.28
3	서구	486,575	27	33.33	합계		1,408,809	81	100.00

대전광역시는 모든 구단위 소지역들이 자체적으로 4개 이상의 충분한 표본조사구 수를 확보하고 있으므로 유사 소지역합성을 고려하지 않고 소지역별로 독립적으로 추정한다.

⑦ 울산광역시

<표5.11> 울산광역시 구/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	중구	233,318	15	23.81	4	북구	123,046	8	12.70
2	남구	345,286	19	30.16	5	울주군	171,685	10	15.87
3	동구	187,043	11	17.46		합계	1,060,378	63	100.00

울산광역시도 모든 구 및 군단위 소지역들이 자체적으로 4개 이상의 충분한 표본조사구 수를 확보하고 있으므로 소지역별로 독립적으로 실업자 총계를 추정한다.

⑧ 경기도

<표5.12> 경기도 시/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	수원시	978,698	16	8.08	17	포천군	148,452	5	2.53
2	안양시	593,967	10	5.05	18	부천시	785,754	13	6.57
3	안산시	598,560	9	4.55	19	광명시	337,175	6	3.03
4	과천시	71,525	3	1.52	20	시흥시	342,351	5	2.53
5	오산시	115,161	4	2.02	21	김포시	183,156	4	2.02
6	군포시	270,326	5	2.53	22	평택시	361,992	8	4.04
7	의왕시	124,772	4	2.02	23	안성시	142,799	5	2.53
8	용인시	455,118	10	5.05	24	고양시	814,493	13	6.57
9	화성시	214,729	6	3.03	25	파주시	226,858	4	2.02
10	성남시	937,780	15	7.58	26	구리시	185,494	4	2.02
11	하남시	124,018	4	2.02	27	남양주시	376,231	11	5.56
12	광주시	154,808	4	2.02	28	가평군	56,211	2	1.01
13	의정부시	368,887	6	3.03	29	이천시	188,367	5	2.53
14	동두천시	75,699	4	2.02	30	여주군	105,084	4	2.02
15	양주군	138,748	4	2.02	31	양평군	82,921	3	1.52
16	연천군	51,902	2	1.01		합계	9,612,036	198	100.00

경기도에서 표본조사구 수가 3개 이하인 지역은 “과천시(3개)”, “연천군(2개)”, “가평군(2)”과 “양평군(3개)”의 4개 지역이다.

<표5.13> 경기도 시/군단위 소지역들의 집락분석 결과

() : 표본조사구 수

구분	집락1	집락2	집락3	집락4	집락5	집락6
시군구	평택시(8), 용인시(10), 파주시(4), 이천시(5), 안성시(5), 광주시(4), 포천군(5), 가평군(2), 양평군(3)	수원시(16), 성남시(15), 안양시(10), 부천시(13), 동두천시(4), 고양시(13), 과천시(3), 구리시(4), 군포시(5), 하남시(4)	연천군(2)	김포시(4), 여주군(4), 화성시(6)	의정부(6), 광명시(6), 의왕시(4), 양주군(4)	안산시(9), 남양주(11), 오산시(4), 시흥시(5)

우선 “가평군(2개)”과 “양평군(3)”은 실업률이 유사한 동일 집락 내의 소지역들이므로 “가평군(2개)+양평군(3)”의 그룹으로 묶어 추정한다. “과천시(3개)”는 동일 집락 내의 “구리시(4개)”의 조사정보를 이용하여 “과천시(3개)+구리시(4개)”그룹으로 묶어 추정한다. “연천군(2개)”은 실업률이 유사한 동일집락 내의 소지역들이 없기 때문에 차선책으로 인근 군지역인 “여주군(4개)”의 조사정보를 이용하여 “연천군(2개)+여주군(4)”그룹으로 묶은 후 실업자 총계를 추정한다. 이러한 경우 “연천군(2개)”의 실업자 총계 추정은 약간의 과소 또는 과대추정이 있을 수 있다.

경기도 지역은 인구의 분포와 경황 특성이 타 지역에 비해 매우 다양한 특성을 나타내기 때문에 소지역추정법을 적용할 경우 편향 값이 매우 크게 추정될 가능성이 있다. 잠재적인 편향 가능성과 추정오차를 줄이기 위해 경기도 지역을 경황 특성이 유사한 4개의 동질적인 그룹으로 구분한 후 각 그룹별로 소지역추정법을 적용하여 경황 특성을 추정한다.

경기 A지구

평택시, 용인시, 파주시, 이천시, 안성시, 광주군, 포천군, 가평군, 양평군

경기 B지구

수원시, 성남시, 안양시, 부천시, 동두천시, 고양시, 과천시, 구리시, 군포시, 하남시

경기 C지구

연천군, 김포시, 여주군, 화성시

경기 D지구

의정부시, 광명시, 의왕시, 양주군, 안산시, 남양주시, 오산시, 시흥시

⑨ 충청북도

<표5.14> 충청북도 시/군단위 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	청주시	594,716	30	37.04	8	음성군	88,783	4	4.94
2	청원군	125,221	6	7.41	9	단양군	38,797	4	4.94
3	진천군	61,483	3	3.70	10	보은군	42,215	2	2.47
4	괴산군	43,220	4	4.94	11	옥천군	59,836	4	4.94
5	증평출장소	31,574	1	1.23	12	영동군	57,319	3	3.70
6	충주시	216,036	12	14.81	합계		1,504,518	81	100.00
7	제천시	145,317	8	9.88					

충청북도에서 표본조사구 수가 3개 이하인 지역은 “진천군(3개)”, “증평출장소(1개)”, “보은군(2)”과 “영동군(3개)”의 4개 지역이다.

<표5.15> 충북 시/군단위 소지역들의 집락분석 결과

(): 표본조사구 수

구분	집락1	집락2	집락3
시군구	청주시(30), 충주시(12), 제천시(8), 청원군(6), 옥천군(4), 진천군(3), 괴산군(4), 음성군(4), 단양군(4)	영동군(3), 증평출장소(1)	보은군(2)

표본조사구 수가 3개 이하인 “진천군(3개)”은 실업률이 유사한 동일집락 내의 “옥천군(4)”의 조사정보를 합성하여 “진천군(3개)+옥천군(4)”그룹으로 묶어 추정한다. “보은군(2개)”의 경우는 실업률이 유사한 동일집락 내의 소지역들이 없기 때문에 인근 유사지역인 “영동군(3개)”과 “증평출장소(1개)”의 소지역들을 같이 묶어 “영동군(3개)+증평출장소(1개)+보은군(2개)”그룹으로 실업자 총계를 추정한다. 이 경우 “영동군(3개)”, “증평출장소(1)”와 “보은군(2)”의 실업자 총계 추정값들은 약간의 과소 또는 과대 추정되는 결과를 보일 수 있다.

⑩ 충청남도

<표5.16> 충남 시/군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	공주시	134,383	7	7.78	10	서천군	72,903	2	2.22
2	논산시	140,793	6	6.67	11	청양군	40,086	2	2.22
3	금산군	63,345	3	3.33	12	홍성군	93,790	4	4.44
4	연기군	81,821	2	2.22	13	서산시	150,504	8	8.89
5	천안시	436,708	26	28.89	14	태안군	67,947	2	2.22
6	아산시	188,372	9	10.00	15	당진군	120,818	5	5.56
7	예산군	100,602	3	3.33	16	계룡출장소	28,883	0	0.00
8	보령시	116,546	7	7.78	합계		1,928,088	90	100.00
9	부여군	90,587	4	4.44					

충청남도에서 표본조사구 수가 3개 이하인 지역은 “금산군(3개)”, “연기군(2개)”, “예산군(3개)”, “서천군(2개)”, “청양군(2개)”, “태안군(2개)”과 “계룡출장소(0개)”의 7개 지역이다. 이들 중 표본조사구 수가 배정되지 않은 “계룡출장소(0개)”는 추정에서 제외한다.

<표5.17> 충남 시/군단위 소지역들의 집락분석 결과
(): 표본조사구 수

구분	집락1	집락2	집락3	집락4
시군구	공주시(7), 금산군(3), 청양군(2)	서천군(2)	보령시(7), 논산시(6), 예산군(3), 태안군(2), 당진군(5), 계룡출장소(0)	천안시(26), 아산시(9), 서산시(8), 연기군(2), 부여군(4), 홍성군(4)

우선 “금산군(3개)”, “청양군(2개)”과 “서천군(2개)”의 3개 소지역들은 “금산군(3개)+청양군(2개)+서천군(2개)”의 그룹으로 묶어 실업자 총계를 추정한다. “예산군(3개)”과 “태안군(2개)”의 2개 지역은 동일집락 내의 소지역들이므로 “예산군(3개)+태안군(2개)”의 그룹으로 묶어 추정한다. “연기군(2개)”은 동일집락 내의 군단위 소지역인 “홍성군(4개)”의 조사정보를 이용하여 “연기군(2개)+홍성군(4개)”의 그룹으로 묶어 추정한다.

⑪ 전라북도

<표5.18> 전라북도 시/군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	전주시	623,897	22	27.16	9	남원시	104,198	4	4.94
2	군산시	277,680	10	12.35	10	임실군	37,548	3	3.70
3	익산시	337,240	11	13.58	11	순창군	34,003	2	2.47
4	김제시	115,683	5	6.17	12	진안군	32,750	2	2.47
5	완주군	87,081	4	4.94	13	무주군	30,385	3	3.70
6	정읍시	152,452	6	7.41	14	장수군	30,521	2	2.47
7	고창군	74,413	3	3.70					
8	부안군	76,072	4	4.94		합계	2,013,923	81	100.00

전라북도에서 표본조사구 수가 3개 이하인 지역은 “고창군(3개)”, “임실군(3개)”, “순창군(2개)”, “진안군(2개)”, “무주군(3개)”과 “장수군(2개)”의 6개 지역이다.

<표5.19> 전라북도 시/군단위 소지역들의 집락분석 결과

(): 표본조사구 수

구분	집락1	집락2	집락3
시군구	정읍시(6), 완주군(4), 진안군(2), 장수군(2), 임실군(3)	무주군(3), 순창군(2), 고창군(3)	전주시(22), 군산시(10), 익산시(11), 남원시(4), 김제시(5), 부안군(4)

“진안군(2개)”, “장수군(2개)”과 “임실군(3)”은 실업률이 유사한 동일집락 내의 소지역들이므로 이들 지역들은 “진안군(2개)+장수군(2개)+임실군(3개)”의 그룹으로 묶어 실업자 총계를 추정한다. 같은 방법으로 “무주군(3개)”, “순창군(2개)”과 “고창군(3개)”의 3개 소지역들은 “무주군(3개)+순창군(2개)+고창군(3개)”의 그룹으로 묶어 추정한다.

⑫ 전라남도

<표5.20> 전라남도 시/군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	담양군	53,672	2	2.22	13	나주시	106,431	5	5.56
2	곡성군	38,930	2	2.22	14	영암군	64,738	2	2.22
3	화순군	79,496	3	3.33	15	함평군	43,667	2	2.22
4	영광군	70,269	3	3.33	16	고흥군	95,960	4	4.44
5	장성군	54,958	2	2.22	17	보성군	59,526	2	2.22
6	목포시	245,666	14	15.56	18	장흥군	51,850	2	2.22
7	무안군	69,178	2	2.22	19	강진군	50,925	2	2.22
8	여수시	320,570	16	17.78	20	해남군	96,590	4	4.44
9	순천시	272,124	13	14.44	21	완도군	65,458	1	1.11
10	광양시	138,468	6	6.67	22	진도군	41,148	1	1.11
11	구례군	33,031	2	2.22	합계		2,104,052	90	100.00
12	신안군	51,361	0	0.00					

전라남도에서 표본조사구 수가 3개 이하인 지역은 “담양군(2개)”, “곡성군(2개)”, “화순군(3개)”, “영광군(3개)”, “장성군(2개)”, “무안군(2개)”, “구례군(2개)”, “신안군(0개)”, “영암군(2개)”, “함평군(2개)”, “보성군(2개)”, “장흥군(2개)”, “강진군(2개)”, “완도군(1개)”과 “진도군(1개)”의 15개 지역이다. 이들 지역 중 표본조사구 수가 배정되지 않은 “신안군(0개)”지역은 추정에서 제외한다.

<표5.21> 전라남도 시/군단위 소지역들의 집락분석 결과
(): 표본조사구 수

구분	집락1	집락2	집락3	집락4	집락5
시군구	목포시(14), 여수시(16), 순천시(13)	광양시(6), 곡성군(2), 구례군(2), 고흥군(4), 보성군(2), 화순군(3), 장흥군(2), 영광군(3), 완도군(1), 진도군(1)	나주시(5), 담양군(2), 해남군(4), 함평군(2), 장성군(2)	강진군(2), 무안군(2)	영암군(2), 신안군(0)

동일집락 내의 “곡성군(2개)”, “구례군(2개)”, “보성군(2개)”, “화순군(3개)”, “장흥군(2개)”, “영광군(3개)”, “완도군(1개)”과 “진도군(1개)”의 8개 지역들은 표본조사구 수를 감안하여 “곡성군(2개)+구례군(2개)+보성군(2개)”의 그룹, “화순군(3개)+장흥군(2개)”의 그룹, “영광군(3개)+완도군(1개)+진도군(1개)”의 그룹으로 묶어 실업자 총계를 추정한다. 같은 방법으로 동일한 집락 내의 “담양군(2개)”, “함평군(2개)”과 “장성군(2개)”의 소지역들을 “담양군(2개)+함평군(2개)+장성군(2개)”의 그룹으로 묶어 추정한다. “영암군(2개)”은 동일집락 내에 실업률이 유사한 소지역들이 없으므로 인근 유사지역들인 “강진군(2개)”과 “무안군(2개)”의 조사정보를 이용하여 “강진군(2개)+무안군(2개)+영암군(2개)”의 그룹으로 묶어 추정한다.

⑬ 경상북도

<표5.22> 경상북도 시/군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	영천시	119,077	4	3.70	13	청송군	33,568	2	1.85
2	경산시	218,638	8	7.41	14	영양군	22,286	2	1.85
3	군위군	34,293	1	0.93	15	봉화군	41,452	3	2.78
4	청도군	51,471	2	1.85	16	김천시	151,764	6	5.56
5	고령군	37,498	2	1.85	17	구미시	348,489	13	12.04
6	성주군	50,933	2	1.85	18	상주시	122,277	6	5.56
7	칠곡군	107,158	3	2.78	19	문경시	89,234	4	3.70
8	포항시	516,576	17	15.74	20	예천군	58,217	2	1.85
9	경주시	288,915	11	10.19	21	영덕군	49,674	2	1.85
10	안동시	182,082	8	7.41	22	울진군	65,878	2	1.85
11	영주시	128,924	5	4.63	23	울릉군	9,950	0	0.00
12	의성군	74,243	3	2.78		합계	2,802,597	108	100.00

경상북도에서 표본조사구 수가 3개 이하인 지역은 “군위군(1개)”, “청도군(2개)”, “고령군(2개)”, “성주군(2개)”, “칠곡군(3개)”, “의성군(3개)”, “청송군(2개)”, “영양군(2개)”, “봉화군(3개)”, “예천군(2개)”, “영덕군(2개)”, “울진군(2개)”과 “울릉군(0개)”의 13개 지역이다. 이들 지역 중 표본조사구 수가 배정되지 않은 “울릉군(0개)”지역은 추정에서 제외한다.

<표5.23> 경상북도 시/군단위 소지역들의 집락분석 결과
(): 표본조사구 수

구분	집락1	집락2	집락3	집락4	집락5
시군구	영주시(5), 상주시(6), 의성군(3), 청송군(2), 영덕군(2), 청도군(2), 고령군(2), 봉화군(3)	포항시(17), 안동시(8), 영천시(4), 문경시(4), 군위군(1), 영양군(2)	경주시(11), 김천시(6), 경산시(8), 성주군(2)	울릉군(0)	구미시(13), 칠곡군(3), 예천군(2), 울진군(2)

동일집락 내에 있는 “의성군(3개)”, “청송군(2개)”, “영덕군(2개)”, “청도군(2개)”, “고령군(2개)”과 “봉화군(3개)”의 6개 지역들은 실업률이 매우 유사한 지역들이므로 표본조사구 수를 고려하여 이들 지역들을 “의성군(3개)+청송군(2개)”, “영덕군(2개)+청도군(2개)”, “고령군(2개)+봉화군(3개)”과 같이 3개의 그룹으로 분할하여 추정한다. 다음으로 “군위군(1개)”, “영양군(2개)”과 “성주군(2개)”을 하나의 그룹인 “군위군(1개)+영양군(2개)+성주군(2개)”으로 묶어 추정한다. 마지막으로 “칠곡군(3개)”, “예천군(2)”과 “울진군(2개)”의 3개 소지역들은 서로 실업률이 유사하므로 “칠곡군(3개)+예천군(2개)+울진군(2개)”의 그룹으로 묶어 실업자 총계를 추정한다.

⑭ 경상남도

<표5.24> 경상남도 시/군단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	창원시	528,152	16	16.16	12	산청군	39,915	2	2.02
2	마산시	434,912	13	13.13	13	통영시	135,845	4	4.04
3	진해시	135,539	4	4.04	14	거제시	180,496	6	6.06
4	김해시	357,149	10	10.10	15	고성군	62,383	3	3.30
5	밀양시	123,393	4	4.04	16	남해군	58,039	2	2.02
6	양산시	202,784	6	6.06	17	하동군	60,370	2	2.02
7	함안군	66,054	2	2.02	18	함양군	45,350	2	2.02
8	창녕군	71,215	2	2.02	19	거창군	68,713	3	3.03
9	진주시	340,669	10	10.10	20	합천군	59,890	2	2.02
10	사천시	119,555	4	4.04					
11	의령군	33,700	2	2.02		합계	3,124,123	99	100.00

경상남도에서 표본조사구 수가 3개 이하인 지역은 “함안군(2개)”, “창녕군(2개)”, “의령군(2개)”, “산청군(2개)”, “고성군(3개)”, “남해군(2개)”, “하동군(2개)”, “함양군(2개)”, “거창군(3개)”과 “합천군(2개)”의 10개 지역이다.

<표5.25> 경상남도 시/군단위 소지역들의 집락분석 결과

() : 표본조사구 수

구분	집락1	집락2	집락3	집락4
시군구	통영시(4), 함안군(2), 하동군(2), 거창군(3)	진주시(10), 사천시(4)	창원시(16), 마산시(13), 진해시(4), 김해시(10), 밀양시(4), 거제시(6), 양산시(6), 의령군(2), 합천군(2)	창녕군(2), 고성군(3), 산청군(2), 함양군(2), 남해군(2)

실업률이 유사한 동일집락 내의 소지역들인 “함안군(2개)”, “하동군(2개)”과 “거창군(3개)”의 3개 소지역들은 “함안군(2개)+하동군(2개)+거창군(3개)”의 그룹으로 묶어 추정한다. “의령군(2개)”과 “합천군(2개)”은 “의령군(2개)+합천군(2개)”으로 묶어 추정한다. 마찬가지로 동일집락 내에 있는 “창녕군(2개)”, “고성군(3개)”, “산청군(2개)”, “함양군(2개)”과 “남해군(2개)”의 5개 소지역들은 표본조사구 수를 감안하여 “창녕군(2개)+고성군(3개)”, “산청군(2개)+함양군(2개)+남해군(2개)”과 같이 2개의 그룹으로 분할하여 실업자 총계를 추정한다.

⑮ 제주도

<표5.26> 제주도 구단위 소지역들의 표본조사구 수 분포

순번	시군구	주민등록 인구	표본조사구		순번	시군구	주민등록 인구	표본조사구	
			조사구 수	구성비 (%)				조사구 수	구성비 (%)
1	제주시	285,097	20	44.44	4	남제주군	77,339	9	20.00
2	서귀포시	85,330	7	15.56	합계	547,964	45	100.00	
3	북제주군	100,208	9	20.00					

제주도는 모든 구단위 소지역들이 자체적으로 4개 이상의 충분한 표본 조사구 수를 확보하고 있으므로 소지역별로 독립적으로 실업자 총계를 추정한다.

5.3 직접추정법

대영역 내에서 소지역 i 에 대한 실업자 총계를 추정하기 위한 직접추정량은 다음과 같은 총계 추정공식이 이용된다.

$$\begin{aligned}\hat{Y}_{i.} &= \sum_{s=1}^2 {}_s\hat{Y}_{i.}, \quad i=1,2,\dots,I; s=1,2; h=1,2,\dots,n_i \\ &= \sum_{s=1}^2 \sum_{h=1}^{n_i} {}_s\hat{Y}_{ih} \\ &= \sum_{s=1}^2 \sum_{h=1}^{n_i} {}_sM_i {}_sY_{ih},\end{aligned}\quad (5.1)$$

여기에서 s 는 성별(남, 여)을 나타내는 첨자, n_i 는 경제활동인구조사에서 소지역 i 의 표본조사구 수, ${}_sY_{ih}$ 는 각 성별에 대해서 소지역 i 의 표본 조사구에서 조사한 실업자 수를 나타낸다. 승수 ${}_sM_i = {}_s\hat{X}_{i.} / {}_sX_{i.}$ 은 $\hat{Y}_{i.}$ 이 불편추정량이 되도록 산정 한다. 여기에서 ${}_s\hat{X}_{i.}$ 은 소지역 i 에 대한 15세 이상의 상주 추계인구를 나타내며, ${}_sX_{i.}$ 은 경제활동인구조사에서 조사된 15세 이상의 조사인구를 나타낸다.

직접추정량 $\hat{Y}_{i.}$ 의 분산은 다음과 같이 주어진다.

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

소지역 i 에 대한 직접추정량 $\hat{Y}_{i.}$ 의 분산에 대한 추정은 통계청의 연속차분산추정공식을 적용하면 다음과 같이 계산된다.

$$\widehat{\text{Var}}(\hat{Y}_{i.}) = \sum_{s=1}^2 {}_sM_i^2 \left(\zeta_i \sum_{h=1}^{n_i} {}_sU_{ih}^2\right) + 2 {}_1M_i {}_2M_i \left(\zeta_i \sum_{h=1}^{n_i} {}_1U_{ih} {}_2U_{ih}\right) \quad (5.2)$$

여기에서

$${}_sU_{ih} = d_s Y_{ih} - {}_s\rho_i \cdot d_s X_{ih},$$

$$d_s Y_{ih} = {}_s Y_{ih} - {}_s Y_{i,h+1} ,$$

$$d_s X_{ih} = {}_s X_{ih} - {}_s X_{i,h+1} ,$$

$$p_i = {}_s Y_{i.} / {}_s X_{i.} ,$$

$$\zeta_i = [1 - n_i / (10N_i)] n_i / [2(n_i - 1)] ,$$

N_i = 소지역 i 에 대한 표본추출틀의 조사구 수.

경제활동인구조사에서 대영역에 포함된 소지역들은 표본설계에 반영된 관심영역이 아니다. 따라서 대영역 표본설계에 기반을 둔 표본 조사로부터 소지역들에 대한 관심 통계량들을 추정한다면 소지역에 할당된 표본 조사구 수가 충분하지 않기 때문에 신뢰할 만한 결과를 얻을 수 없게 된다. 이러한 관점에서 현행 경제활동인구조사 자료로부터 소지역 추정값들의 신뢰성을 확보하기 위한 다음과 같은 합성추정법 및 복합추정법이 제안될 수 있다.

5.4 합성 추정법

대영역 표본설계에 기반을 둔 통계청의 직접추정량은 각 소지역에 할당된 표본 조사구의 수가 충분하지 않기 때문에 소지역 실업통계의 정확도를 제공하지는 못한다. 소지역 i 에 대한 합성추정량 \hat{Y}_i^s 는 소지역 i 와 특성이 유사한 인근 지역의 정보를 추정에 이용하는 간접적인 설계 기반 추정량이다.

우선 대영역 내에 I 개의 시군구 단위의 소지역들이 있다고 가정하자. "Borrow Strength"를 적용하기 위해 대영역을 특성이 유사한 시단위, 군단위 및 구단위들의 3개의 그룹으로 분할하고, 각 그룹들을 4개의 성별(남, 여)-연령대별(15-29세, 30세이상) 범주로 구분한다. 여기에서 $I = \sum_{k=1}^3 I_k$ 이며, 시, 군 및 구 그룹들은 각각 I_1 , I_2 , I_3 개의 동질적인 소지역 단위들로 구성된다.

"Borrow Strength"를 적용하여 정의되는 소지역 i 에 대한 합성추정량

\hat{Y}_i^S 는 해당 소지역과 유사한 정보를 갖는 인근 소지역들의 정보를 이용하여 추정되므로 추정오차는 직접추정량에 비해 현저하게 줄어들 수 있으나 해당 소지역과 인근 유사지역의 정보가 동질적이지 못할 경우 편향이 발생할 가능성이 있다.

합성추정량을 정의하기 위해 다음과 같은 기호들이 이용되었다.

N_i = 표본추출틀에서 소지역 i 의 조사구 수,

n_i = 경제활동인구조사에서 소지역 i 에 할당된 표본조사구 수,

${}_jP_{t,i}^C$ = t 년 센서스로부터 추계된 j 범주에 대한 소지역 i 의 상주인구,

${}_jP_{t,i}^R$ = j 범주에 대한 소지역 i 의 t 년 주민등록인구,

${}_jP_{month,i}^R$ = j 범주에 대한 소지역 i 의 경제활동인구조사 달의 주민등록 인구,

${}_j\hat{X}_i$ = j 범주에 대한 소지역 i 의 상주추정인구,

${}_jY_{ih}$ = j 범주에 대한 소지역 i 의 h 번째 표본조사구의 실업자 수.

I_k 개의 소지역들을 포함하고 있는 부차관심영역들인 각 시, 군, 구 그룹들을 특성 기준에 따라 유사성을 갖는 4개의 성별(남, 여)-연령대별(15-29세, 30세이상) 범주들로 구분할 때 각 그룹 내에서 소지역 i 의 실업자 총계에 대한 합성추정량 \hat{Y}_i^S 는 다음과 같이 주어질 수 있다.

$$\hat{Y}_i^S = \sum_{j=1}^4 \frac{{}_j\hat{P}_i}{{}_j\hat{X}_i} {}_j\hat{Y}_{dt} , \quad i = 1, 2, \dots, I_k , \quad (5.4)$$

여기에서

$${}_j\hat{P}_i = \frac{{}_jP_{t,i}^C}{{}_jP_{t,i}^R} {}_jP_{month,i}^R ,$$

$${}_j\hat{X}_i = \sum_{i=1}^k {}_j\hat{X}_i ,$$

$${}_j\hat{Y}_{dir} = \sum_{i=1}^k \sum_{h=1}^{n_j} {}_jM_{i,j} Y_{ih}$$

로 주어진다. ${}_j\hat{P}_i$ 는 행정보고자료로부터 산정된 j 범주에 대한 소지역 i 의 상주추정인구를 나타내며, ${}_j\hat{X}_i$ 는 경제활동인구조사 자료로부터 산정되는 j 범주에 대한 상주추정인구를 나타낸다. 또한 ${}_j\hat{Y}_{dir}$ 는 j 번째 성별-연령대 별 범주의 실업자 총계에 대한 직접추정량을 의미하며, 경제활동인구조사 자료로부터 산정된다. 소지역 i 의 j 범주에 대한 승수는 ${}_jM_{i,j} = {}_j\hat{X}_i / X_i$ 로 주어진다.

합성추정량 \hat{Y}_i^s 에 대한 정확도의 측도로써 다음과 같은 합성추정량의 평균제곱오차를 고려할 수 있다.

$$MSE(\hat{Y}_i^s) = Var(\hat{Y}_i^s) + [Bias(\hat{Y}_i^s)]^2$$

위 식에서 $Var(\hat{Y}_i^s)$ 는 ${}_j\hat{P}_i / {}_j\hat{X}_i = const$ 를 가정한다면 다음 식과 같이 주어질 수 있다.

$$\begin{aligned} Var(\hat{Y}_i^s) &= \sum_{j=1}^J \left(\frac{{}_j\hat{P}_i}{{}_j\hat{X}_i} \right)^2 Var({}_j\hat{Y}_{dir}) \\ &+ 2 \sum_{j < l} \left(\frac{{}_j\hat{P}_i}{{}_j\hat{X}_i} \right) \left(\frac{{}_l\hat{P}_i}{{}_l\hat{X}_i} \right) Cov({}_j\hat{Y}_{dir}, {}_l\hat{Y}_{dir}) \end{aligned} \quad (5.5)$$

j 범주의 실업자 총계에 대한 직접추정량은

$${}_j\hat{Y}_{dir} = \sum_{i=1}^k \sum_{h=1}^{n_j} M_{j,i,h} Y_{ijh}$$

로 주어지므로 직접추정량 ${}_j\hat{Y}_{dir}$ 의 분산과 공분산은 연속차 분산추정방법에 의해 다음 식으로부터 추정될 수 있다.

$$\widehat{Var}({}_j\hat{Y}_{dir}) = M_j^2 \zeta_j \sum_{i=1}^k \sum_{h=1}^{n_j} U_{ijh}^2$$

$$\widehat{Cov}(\hat{Y}_{dir}, \hat{Y}_{dir}) = M_j M_l \zeta_j \sum_{i=1}^k \sum_{h=1}^{n_j} U_{ijh} U_{ilh} ,$$

여기에서

$$U_{ijh} = d_j Y_{ijh} - \frac{Y_{\cdot j \cdot}}{X_{\cdot j \cdot}} \cdot d_j X_{ijh} ,$$

$$d_j Y_{ijh} = Y_{ijh} - Y_{ij, h+1} ,$$

$$d_j X_{ijh} = X_{ijh} - X_{ij, h+1} ,$$

$$\zeta_j = (1 - f_j) n_j / [2(n_j - 1)] ,$$

$$f_j = n_j / (10N_j)$$

로 주어진다.

따라서 합성추정량 \hat{Y}_i^s 의 추정분산은 다음과 같이 주어진다.

$$\begin{aligned} \widehat{Var}(\hat{Y}_i^s) &= \left(\frac{j \hat{P}_i}{j \hat{X}_{\cdot}} \right)^2 \left(M_j^2 \zeta_j \sum_{i=1}^k \sum_{h=1}^{n_j} U_{ijh}^2 \right) \\ &+ 2 \sum_{j < l} \left(\frac{j \hat{P}_i}{j \hat{X}_{\cdot}} \right) \left(\frac{l \hat{P}_i}{l \hat{X}_{\cdot}} \right) \left(M_j M_l \zeta_j \sum_{i=1}^k \sum_{h=1}^{n_j} U_{ijh} U_{ilh} \right) \end{aligned} \quad (5.6)$$

소지역 i 에 대한 합성추정량 \hat{Y}_i^s 의 추정분산은 위의 (5.6)식과 같이 명시된 절차에 의해 계산될 수 있으나, 현행 경제활동인구조사 체계에서 편향에 대한 추정은 결코 쉬운 문제가 아니다. 소지역 i 의 실업자 총계에 대한 참값을 센서스 자료를 이용하여 결정할 수는 있으나 시점 상으로 서로 상이한 양상을 보일 가능성이 있기 때문에 편향 추정에 직접적으로 이용될 수는 없다.

이러한 문제점에 기인하여 Ghosh and Rao (1994)는 $Cov(\hat{Y}_i, \hat{Y}_i^s) = 0$ 의 가정 하에서 합성추정량 \hat{Y}_i^s 의 평균제곱오차를 추정할 경우 다음과 같은 근사적인 불편추정량을 이용할 것을 제안한 바 있다.

$$mse(\hat{Y}_{i.}^s) \approx (\hat{Y}_{i.}^s - \hat{Y}_{i.})^2 - \widehat{Var}(\hat{Y}_{i.}).$$

그러나 위의 추정량은 소지역에 배정된 표본조사구 수가 충분하지 못할 경우 직접추정값의 불안정에 기인하여 합성추정값의 평균제곱오차의 추정값이 음의 값이 나올 가능성도 있다. 따라서 소지역에 배정된 표본조사구 수가 충분하지 못한 한국의 경찰조사 체계에 적용하기에는 무리가 있는 추정공식이다. 또 다른 추정량의 정확도에 대한 측도로써 대영역 내의 모든 소지역들에 대해 평균제곱오차 추정값들의 평균을 취하는 방법이 이용될 수 있으나 이 측도는 전자보다는 좀 더 안정적일 수는 있지만 각 소지역에 대한 평균제곱오차 추정값들을 제공하지는 못한다.

위에서 언급된 사실들을 보완할 수 있는 하나의 대안으로써 잭나이프 추정 방법이 고려될 수 있다. 잭나이프 추정방법의 첫 번째 단계는 경찰조사 자료로부터 반복표본을 생성하는 것이다. 우선 소지역 i 내에서 하나의 표본조사구가 교대로 선택되어 표본으로부터 제거된 후 나머지 표본조사구들에 대해서 승수가 보정된다. 반복표본들은 조사구의 갯수 만큼 생성되며, 이 들 반복표본들을 이용하여 새로운 합성추정값들이 다시 계산된다. 구체적인 절차를 설명하면 다음과 같다.

(I) 대영역 내에서 분할된 유사성질을 갖는 부차그룹(시, 군 및 구 그룹)에 대해서 소지역 i 로부터 h 번째 표본조사구를 제거한 후, 실업자 조사 총계 Y 에 대한 다음과 같은 반복표본을 생성한다.

$$S_{i(h)} = \{Y_{11}, \dots, Y_{1n_i}; \dots; Y_{i1}, \dots, Y_{i,h-1}, Y_{i,h+1}, \dots, Y_{in_i}; Y_{i1}, \dots, Y_{in_i}\}.$$

기호 $i(h)$ 는 새로운 합성추정값을 계산하기 위해 소지역 i 로부터 h 번째 표본조사구가 제거되었다는 것을 나타낸다. 소지역 i 에 대해서 총 n_i 개의 반복표본이 생성되며, k 번째 부차그룹에 대한 반복표본의 개수는 총

$$n = \sum_{i=1}^L n_i \text{ 개이다.}$$

(II) 주어진 소지역 i 에서 해당 소지역 내에 남아있는 $n_i - 1$ 개 표본조사구의 전체 조사가구들에 대해서 승수에 대한 보정이 이루어진다. 보정

된 승수값은 다음과 같다.

$${}_jM_i^{adj} = \frac{n_i}{n_i - 1} {}_jM_i, \quad j = 1, 2, 3, 4$$

(III) k 번째 부차그룹 내에 남아 있는 $n-1$ 개 표본조사구들을 이용하여 소지역 i 에 대한 새로운 합성추정값 $\hat{Y}_{i.}^S(h)$ 를 계산한다.

위의 절차는 해당 부차관심영역 내에 있는 모든 표본조사구들에 대해서 반복되며, 이로부터 n 개의 서로 다른 실업자 총계에 대한 합성추정값들이 생성된다. 소지역 i 에 대해서는 n_i 개의 서로 다른 합성추정값들이 얻어진다.

소지역 i 의 실업자 총계에 대한 잭나이프 평균제곱오차 추정식은 다음과 같이 주어진다.

$$mse_j(\hat{Y}_{i.}^S) = \widehat{Var}_j(\hat{Y}_{i.}^S) + [\widehat{Bias}_j(\hat{Y}_{i.}^S)]^2 \quad (5.7)$$

여기에서

$$\widehat{Var}_j(\hat{Y}_{i.}^S) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} [\hat{Y}_{i.}^S(h) - \frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}^S]^2, \quad (5.8)$$

$$\widehat{Bias}_j(\hat{Y}_{i.}^S) = (n_i - 1) \left[\frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}^S(h) - \hat{Y}_{i.}^S \right] \quad (5.9)$$

로 주어진다.

5.5 복합 추정법

경제활동인구조사 자료로부터 직접적으로 추정된 소지역 i 에 대한 직접추정량 $\hat{Y}_{i.}$ 은 해당 소지역에 할당된 표본 수가 충분하지 않기 때문에 추정값들의 신뢰성을 확보할 수 없다. 또한 인근 지역의 유사정보를 이용하여 추정되는 합성추정량 $\hat{Y}_{i.}^S$ 는 잠재적인 편향 가능성이 항상 내재되어 있다. 따라서 소지역에 배정된 표본 수가 적을 경우, 표본조사 자료만을

이용하여 추정되는 직접추정량의 불안정성과 합성추정량의 편향 가능성을 보완하기 위해 직접추정량 $\hat{Y}_{i.}$ 와 합성추정량 $\hat{Y}_{i.}^s$ 의 가중평균을 이용한 다음과 같은 복합추정량 $\hat{Y}_{i.}^c$ 가 고려될 수 있다.

$$\hat{Y}_{i.}^c = w_i \hat{Y}_{i.} + (1 - w_i) \hat{Y}_{i.}^s, \quad i = 1, 2, \dots, I_k \quad (5.10)$$

여기에서 가중치 w_i 는 0과 1사이의 값을 취한다.

이 때 복합추정량 $\hat{Y}_{i.}^c$ 의 평균제곱오차는 다음 식과 같이 주어진다.

$$\begin{aligned} MSE(\hat{Y}_{i.}^c) &= w_i^2 MSE(\hat{Y}_{i.}) + (1 - w_i)^2 MSE(\hat{Y}_{i.}^s) \\ &\quad + 2w_i(1 - w_i) E(\hat{Y}_{i.} - Y_{i.}^*)(\hat{Y}_{i.}^s - Y_{i.}^*) \end{aligned} \quad (5.11)$$

여기에서 $Y_{i.}^*$ 는 소지역 i 의 실업자 총계에 대한 참값을 나타낸다. 위의 (5.11)식을 w_i 의 함수로 가정하여 가중치 w_i 에 대해서 미분하면, 평균제곱오차를 최소화하는 다음과 같은 가중치를 산정할 수 있다.

$$w_{i(opt)}^* = \frac{MSE(\hat{Y}_{i.}^s) - E(\hat{Y}_{i.} - Y_{i.}^*)(\hat{Y}_{i.}^s - Y_{i.}^*)}{MSE(\hat{Y}_{i.}^s) + MSE(\hat{Y}_{i.}) - 2E(\hat{Y}_{i.} - Y_{i.}^*)(\hat{Y}_{i.}^s - Y_{i.}^*)} \quad (5.12)$$

여기에서 직접추정량 $\hat{Y}_{i.}$ 은 불편추정량이 되도록 산정하므로 (5.12)식에서 평균제곱오차 $MSE(\hat{Y}_{i.})$ 은 $Var(\hat{Y}_{i.})$ 과 동일한 값을 갖는다. 또한 (5.12)식에서 $Cov(\hat{Y}_{i.}, \hat{Y}_{i.}^s) = 0$ 를 가정한다면, 가중치 $w_{i(opt)}^*$ 는 다음 식으로 근사될 수 있다.

$$w_{i(opt)} = \frac{MSE(\hat{Y}_{i.}^s)}{MSE(\hat{Y}_{i.}^s) + Var(\hat{Y}_{i.})} \quad (5.13)$$

위의 (5.13)식에서 최적가중치 $w_{i(opt)}$ 는 경제활동인구조사 자료로부터 추정되어야 할 값이다. $MSE(\hat{Y}_{i.}^s)$ 에 대해서는 (5.7)식에서 주어진 $mse_j(\hat{Y}_{i.}^s)$ 으로, $Var(\hat{Y}_{i.})$ 에 대해서는 (5.2)식에서 주어진 $\widehat{Var}(\hat{Y}_{i.})$ 으로 추정될 수 있고, 이때 최적가중치에 대한 추정식은 다음과 같이 주어진다.

$$\hat{w}_{i(opt)} = \frac{mse_J(\hat{Y}_{i.}^s)}{mse_J(\hat{Y}_{i.}^s) + \widehat{Var}(\hat{Y}_{i.})} \quad (5.14)$$

따라서 (5.10)식의 복합추정량은 경제활동인구조사 자료로부터 추정된 최적가중치 $\hat{w}_{i(opt)}$ 를 이용하여 다음 식으로부터 추정될 수 있다.

$$\hat{Y}_{i.}^c = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s \quad (5.15)$$

소지역 i 에 대한 복합추정량의 평균제곱오차 $MSE(\hat{Y}_{i.}^c)$ 는 경제활동인구조사 자료로부터 추정되어야 하며 좀 더 안정적인 평균제곱오차에 대한 추정값들을 산출하기 위해 잭나이프 추정방법이 적용될 수 있다. 소지역 i 에서 실업자 총계에 대한 복합추정값들의 잭나이프 평균제곱오차는 다음 추정식으로부터 산정할 수 있다.

$$mse_J(\hat{Y}_{i.}^c) = \widehat{Var}(\hat{Y}_{i.}^c) + [\widehat{Bias}_J(\hat{Y}_{i.}^c)]^2 \quad (5.16)$$

여기에서

$$\widehat{Var}_J(\hat{Y}_{i.}^c) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} \left[\hat{Y}_{i.}^c(h) - \frac{1}{n_i} \sum_{l=1}^{n_i} \hat{Y}_{i.}^c(l) \right]^2, \quad (5.17)$$

$$\widehat{Bias}_J(\hat{Y}_{i.}^c) = (n_i - 1) \left[\frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}^c(h) - \hat{Y}_{i.}^c \right], \quad (5.18)$$

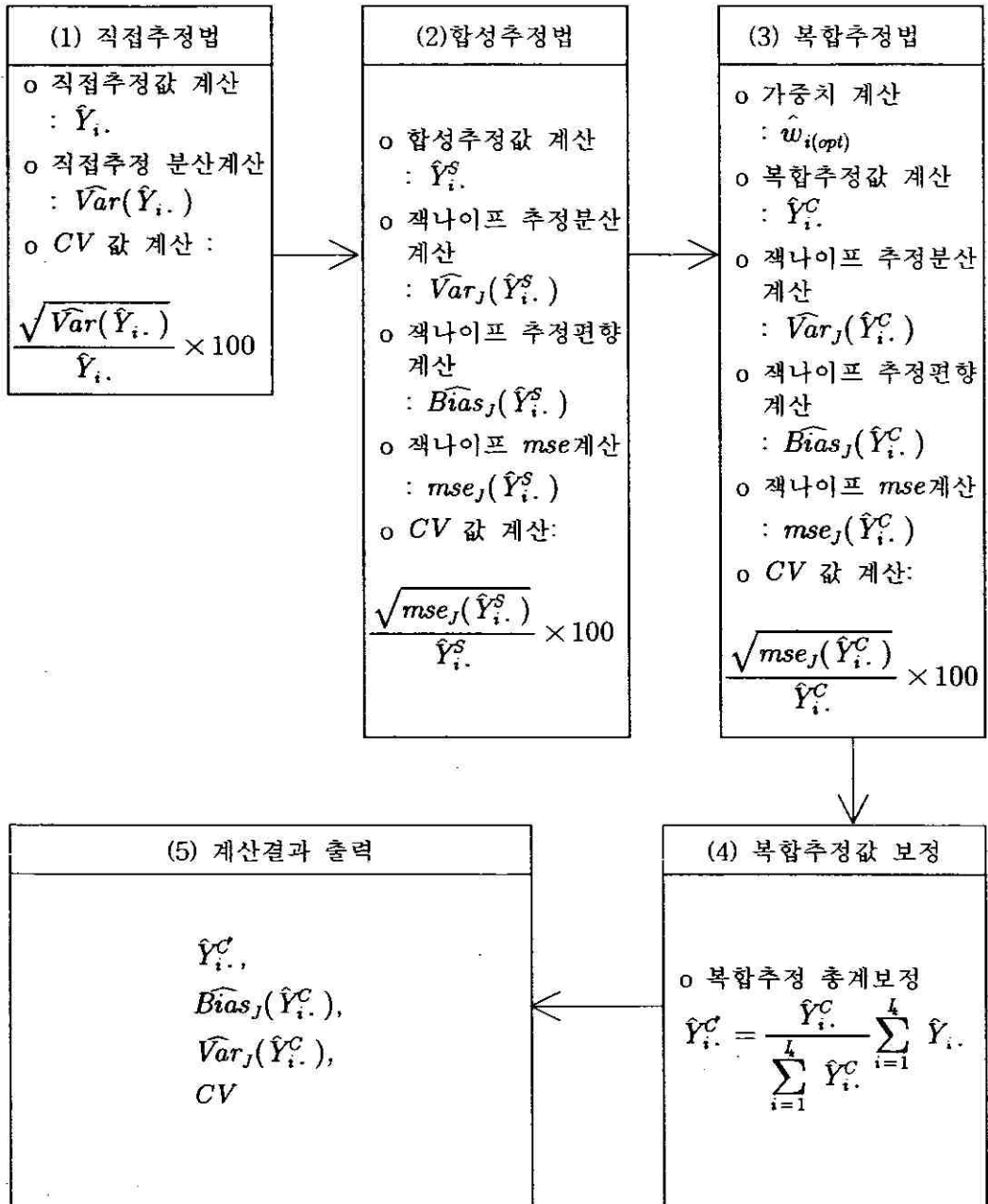
$$\hat{Y}_{i.}^c(h) = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s(h) \quad (5.19)$$

로 주어진다.

5.6 SAS 프로그램 알고리즘

5.6.1 프로그램 순서도

SAS 프로그램을 순서도로 나타내면 다음과 같다.



5.6.2 세부 알고리즘

대영역인 충청북도는 3개의 시 단위 소지역(청주, 충주, 제천)들과, 8개의 군단위 소지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)들로 구성되어있다. 충청북도의 경제활동인구조사 자료에 대한 시군구 단위 소지역 추정을 예로 설명한다.

시군 단위 소지역들의 지역코드는 “청주=1”, “충주=2”, “제천=3”, “보은=4”, “옥천=5”, “영동=6”, “괴산=7”, “진천=8”, “음성=9”, “청원=10”, “단양=11”로 나타내고, 각 시군 단위 소지역들에 대한 표본조사구의 수는 “청주= n_1 ”, “충주= n_2 ”, “제천= n_3 ”, “보은= n_4 ”, “옥천= n_5 ”, “영동= n_6 ”, “괴산= n_7 ”, “진천= n_8 ”, “음성= n_9 ”, “청원= n_{10} ”, “단양= n_{11} ”로 나타내기로 한다.

(1) 직접추정법

(Step1) 직접추정값 계산($\hat{Y}_{i.}$)

대영역 내의 I 개의 소지역에 대한 실업자 총계 추정값은 경제활동인구조사 자료에 근거하여 다음 직접 추정 공식을 이용하여 계산한다. 경제활동인구조사자료는 조사구별로 자료가 정리되어 있고, 소지역 i 의 자료는 동부와 읍면부 단위의 조사구들로 구성되어 있다.

충청북도 내의 시군 단위 소지역들을 예로 설명하면, 소지역 i 가 청주시와 같이 시 단위 지역인 경우는 동단위의 조사구들로 구성되어 있고, 청원군과 같이 군 단위 지역인 경우 읍면 단위의 조사구들로 구성되어 있으나, 충주시의 경우에는 동단위와 읍면단위 조사구가 혼합되어 있다.

따라서 위의 3가지 경우를 고려하여 대영역인 하나의 집락 내에서 해당 소지역들에 대한 실업자 총계 추정 프로그램을 작성하여야 한다

$$\begin{aligned}\hat{Y}_{i.} &= \sum_{s=1}^2 s \hat{Y}_{i.s}, \quad i=1,2,\dots,I; s=1,2; h=1,2,\dots,n_i \\ &= \sum_{s=1}^2 \sum_{h=1}^{n_i} s \hat{Y}_{sh}\end{aligned}$$

$$= \sum_{s=1}^2 \sum_{h=1}^{n_i} {}_sM_i {}_sY_{ih} ,$$

여기에서

s = 성별(남1, 여2)을 나타내는 첨자,

n_i = 경제활동인구조사에서 소지역 i 의 표본조사구 수,

${}_sY_{ih}$ = 각 성별에 대해서 소지역 i 의 표본조사구에서 조사한 실업자수,

${}_sM_i = \hat{X}_{i.} / {}_sX_{i.}$ (주어진 값:데이터 필드에 포함됨),

$\hat{X}_{i.}$ = 소지역 i 에 대한 15세 이상의 상주 추계인구,

${}_sX_{i.}$ = 경제활동인구조사에서 조사된 15세 이상의 조사인구.

※ 소지역 i 의 표본조사구들이 동단위, 읍면단위, 또는 동과 읍면단위의 혼합으로 구성되어 있어도 일반적으로 적용가능

(Step2) 직접추정값의 분산 계산

소지역 i 에 대한 직접 추정값의 분산은 다음 식을 적용하여 추정한다.

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

여기에서

$${}_sU_{ih} = d_s Y_{ih} - \rho_i \cdot d_s X_{ih} ,$$

$$d_s Y_{ih} = {}_sY_{ih} - {}_sY_{i,h+1} ,$$

$$d_s X_{ih} = {}_sX_{ih} - {}_sX_{i,h+1} ,$$

$$\rho_i = {}_sY_{i.} / {}_sX_{i.} ,$$

$$\zeta_i \approx n_i / [2(n_i - 1)] ,$$

N_i = 소지역 i 에 대한 표본추출들의 조사구 수.

※ 소지역 i 가 동단위 표본조사구들만으로 구성되어 있을 경우와, 소지역 i 가 읍면단위 표본조사구들만으로 구성되어 있을 경우는 위의 분산 추정공식을 이용하여 계산하고,

※ 소지역 i 가 동단위와 읍면단위의 표본조사구들로 혼합되어 있을 경우에는(예를 들면 충북지역에서 충주시와 같은 경우임), 동단위 표본조사구들만으로 추정분산을 계산하고, 읍면단위 표본조사구들만으로 추정분산을 계산한 후, 이들을 합하여 소지역 i 의 추정분산을 산출함

(Step3) 직접추정값들의 CV 값 계산

$$CV(\hat{Y}_{i.}) = \frac{\sqrt{\widehat{Var}(\hat{Y}_{i.})}}{\hat{Y}_{i.}} \times 100, \quad i = 1, 2, \dots, 11$$

(Step4) 이상의 직접추정법에 의한 계산 결과를 저장

$$\hat{Y}_{1.}, \hat{Y}_{2.}, \dots, \hat{Y}_{11.},$$

$$\widehat{Var}(\hat{Y}_{1.}), \widehat{Var}(\hat{Y}_{2.}), \dots, \widehat{Var}(\hat{Y}_{11.})$$

$$CV(\hat{Y}_{1.}), CV(\hat{Y}_{2.}), \dots, CV(\hat{Y}_{11.})$$

(2) 합성추정법

(Step1) 합성추정값을 계산하기 위한 사전작업으로써 대영역인 충북지역을 시지역(청주, 충주, 제천), 군지역(보은, 옥천, 괴산, 진천, 영동, 음성, 청원, 단양), 구지역(충북에는 구지역이 없음)으로 데이터 셀을 분할하고 표본조사구별로 경제활동인구조사 자료를 작성한다. 이 후 각각의 그룹에 대해서 다음을 계산한다.

① 범주별 상주추계인구 자료 작성(통계청 추계자료 이용)

○ 시지역(청주, 충주, 제천)

청주A1 = 청주 1범주 상주추계인구, 청주A2 = 청주 2범주 상주추계인구,

청주 A3 = 청주 3범주 상주추계인구, 청주 A4 = 청주 4범주 상주추계인구,

충주 A1 = 충주 1범주 상주추계인구, 충주 A2 = 충주 2범주 상주추계인구,

충주 A3 = 충주 3범주 상주추계인구, 충주 A4 = 충주 4범주 상주추계인구,

제천 A1 = 제천 1범주 상주추계인구, 제천 A2 = 제천 2범주 상주추계인구,

제천 A3 = 제천 3범주 상주추계인구, 제천 A4 = 제천 4범주 상주추계인구,

○ 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)

보은 A1 = 보은 1범주 상주추계인구, 보은 A2 = 보은 2범주 상주추계인구,

보은 A3 = 보은 3범주 상주추계인구, 보은 A4 = 보은 4범주 상주추계인구,

옥천 A1 = 옥천 1범주 상주추계인구, 옥천 A2 = 옥천 2범주 상주추계인구,

옥천 A3 = 옥천 3범주 상주추계인구, 옥천 A4 = 옥천 4범주 상주추계인구,

.....

단양 A1 = 단양 1범주 상주추계인구, 단양 A2 = 단양 2범주 상주추계인구,

단양 A3 = 단양 3범주 상주추계인구, 단양 A4 = 단양 4범주 상주추계인구.

② 범주별 상주조사인구 카운트(경찰조사에서 단순히 카운트된 수)

○ 시지역(청주, 충주, 제천)

청주 B1 = 청주 1범주 상주조사인구, 청주 B2 = 청주 2범주 상주조사인구,

청주 B3 = 청주 3범주 상주조사인구, 청주 B4 = 청주 4범주 상주조사인구,

충주 B1 = 충주 1범주 상주조사인구, 충주 B2 = 충주 2범주 상주조사인구,

충주 B3 = 충주 3범주 상주조사인구, 충주 B4 = 충주 4범주 상주조사인구,

제천B1 =제천 1범주 상주조사인구, 제천B2 =제천 2범주 상주조사인구,
제천B3 =제천 3범주 상주조사인구, 제천B4 =제천 4범주 상주조사인구,

○ 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)

보은B1 =보은 1범주 상주조사인구, 보은B2 =보은 2범주 상주조사인구,
보은B3 =보은 3범주 상주조사인구, 보은B4 =보은 4범주 상주조사인구,

.....

단양B1 =단양 1범주 상주조사인구, 단양B2 =단양 2범주 상주조사인구,
단양B3 =단양 3범주 상주조사인구, 단양B4 =단양 4범주 상주조사인구.

③ 범주별 실업자 총계 카운트(경찰조사에서 단순히 카운트된 수)

○ 시지역(청주, 충주, 제천)

청주C1 =청주 1범주 실업자 총계, 청주C2 =청주 2범주 실업자 총계,
청주C3 =청주 3범주 실업자 총계, 청주C4 =청주 4범주 실업자 총계,

충주C1 =충주 1범주 실업자 총계, 충주C2 =충주 2범주 실업자 총계,
충주C3 =충주 3범주 실업자 총계, 충주C4 =충주 4범주 실업자 총계,

제천C1 =제천 1범주 실업자 총계, 제천C2 =제천 2범주 실업자 총계,
제천C3 =제천 3범주 실업자 총계, 제천C4 =제천 4범주 실업자 총계,

○ 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)

보은C1 =보은 1범주 실업자 총계, 보은C2 =보은 2범주 실업자 총계,
보은C3 =보은 3범주 실업자 총계, 보은C4 =보은 4범주 실업자 총계,

.....

단양C1 = 단양 1범주 실업자 총계, 단양C2 = 단양 2범주 실업자 총계,
 단양C3 = 단양 3범주 실업자 총계, 단양C4 = 단양 4범주 실업자 총계,

④ 범주별 경제활동인구 총계 카운트(경찰조사에서 단순히 카운된 수)

o 시지역(청주, 충주, 제천)

청주D1 = 청주 1범주 경찰인구총계, 청주D2 = 청주 2범주 경찰인구총계,
 청주D3 = 청주 3범주 경찰인구총계, 청주D4 = 청주 4범주 경찰인구총계,

충주D1 = 충주 1범주 경찰인구총계, 충주D2 = 충주 2범주 경찰인구총계,
 충주D3 = 충주 3범주 경찰인구총계, 충주D4 = 충주 4범주 경찰인구총계,

제천D1 = 제천 1범주 경찰인구총계, 제천D2 = 제천 2범주 경찰인구총계,
 제천D3 = 제천 3범주 경찰인구총계, 제천D4 = 제천 4범주 경찰인구총계,

o 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)

보은D1 = 보은 1범주 경찰인구총계, 보은D2 = 보은 2범주 경찰인구총계,
 보은D3 = 보은 3범주 경찰인구총계, 보은D4 = 보은 4범주 경찰인구총계,

.....

단양D1 = 단양 1범주 경찰인구총계, 단양D2 = 단양 2범주 경찰인구총계,
 단양D3 = 단양 3범주 경찰인구총계, 단양D4 = 단양 4범주 경찰인구총계,

(Step2) 합성추정값 계산

① 시지역 합성추정값 계산

o 범주별 실업자 총계 계산

$$\begin{aligned}
 {}_1\hat{Y}_{\#} &= (\text{청주}A1/\text{청주}B1) * \text{청주}C1 \\
 &\quad + (\text{충주}A1/\text{충주}B1) * \text{충주}C1 \\
 &\quad + (\text{제천}A1/\text{제천}B1) * \text{제천}C1
 \end{aligned}$$

$$\begin{aligned}
{}_2\hat{Y}_{dir} &= (\text{청주}A2/\text{청주}B2)*\text{청주}C2 \\
&+ (\text{충주}A2/\text{충주}B2)*\text{충주}C2 \\
&+ (\text{제천}A2/\text{제천}B2)*\text{제천}C2
\end{aligned}$$

$$\begin{aligned}
{}_3\hat{Y}_{dir} &= (\text{청주}A3/\text{청주}B3)*\text{청주}C3 \\
&+ (\text{충주}A3/\text{충주}B3)*\text{충주}C3 \\
&+ (\text{제천}A3/\text{제천}B3)*\text{제천}C3
\end{aligned}$$

$$\begin{aligned}
{}_4\hat{Y}_{dir} &= (\text{청주}A4/\text{청주}B4)*\text{청주}C4 \\
&+ (\text{충주}A4/\text{충주}B4)*\text{충주}C4 \\
&+ (\text{제천}A4/\text{제천}B4)*\text{제천}C4
\end{aligned}$$

○ 실업자 총계에 대한 합성추정값 계산

-청주시

$$\begin{aligned}
\hat{Y}_1^s &= (\text{청주}A1*_1\hat{Y}_{dir})/(\text{청주}A1+\text{충주}A1+\text{제천}A1) \\
&+ (\text{청주}A2*_2\hat{Y}_{dir})/(\text{청주}A2+\text{충주}A2+\text{제천}A2) \\
&+ (\text{청주}A3*_3\hat{Y}_{dir})/(\text{청주}A3+\text{충주}A3+\text{제천}A3) \\
&+ (\text{청주}A4*_4\hat{Y}_{dir})/(\text{청주}A4+\text{충주}A4+\text{제천}A4)
\end{aligned}$$

-충주시

$$\begin{aligned}
\hat{Y}_2^s &= (\text{충주}A1*_1\hat{Y}_{dir})/(\text{청주}A1+\text{충주}A1+\text{제천}A1) \\
&+ (\text{충주}A2*_2\hat{Y}_{dir})/(\text{청주}A2+\text{충주}A2+\text{제천}A2) \\
&+ (\text{충주}A3*_3\hat{Y}_{dir})/(\text{청주}A3+\text{충주}A3+\text{제천}A3) \\
&+ (\text{충주}A4*_4\hat{Y}_{dir})/(\text{청주}A4+\text{충주}A4+\text{제천}A4)
\end{aligned}$$

-제천시

$$\begin{aligned}
\hat{Y}_3^s &= (\text{제천}A1*_1\hat{Y}_{dir})/(\text{청주}A1+\text{충주}A1+\text{제천}A1) \\
&+ (\text{제천}A2*_2\hat{Y}_{dir})/(\text{청주}A2+\text{충주}A2+\text{제천}A2)
\end{aligned}$$

$$+(\text{제천}A3*_3 \hat{Y}_{dir})/(\text{청주}A3+\text{충주}A3+\text{제천}A3)$$

$$+(\text{제천}A4*_4 \hat{Y}_{dir})/(\text{청주}A4+\text{충주}A4+\text{제천}A4)$$

(output) 합성추정값 계산결과 저장

$$\hat{Y}_1^s, \hat{Y}_2^s, \hat{Y}_3^s.$$

② 군지역 합성추정값 계산

위의 ①과 같은 방법으로 계산함

(Step3) 시지역(청주, 충주, 제천)에서 반복 합성추정량 생성

① 청주시에서 반복 합성추정량들을 생성

다음의 과정을 반복하면 청주시의 표본조사구 수(= n_1) 만큼 반복 합성추정량들이 생성됨. 청주시에서 l 번째 표본조사구를 제외하고 나머지 표본조사구들을 이용하여 새로운 합성추정값을 다음 절차에 의해 반복적으로 생성함(= $\hat{Y}_1^s(l)$, $l=1,2,\dots,n_1$)

(절차1) 청주시 범주별 상주조사인구 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

$$\text{청주}B1(l) = \text{청주시 1범주 상주조사인구,}$$

$$\text{청주}B2(l) = \text{청주시 2범주 상주조사인구,}$$

$$\text{청주}B3(l) = \text{청주시 3범주 상주조사인구,}$$

$$\text{청주}B4(l) = \text{청주시 4범주 상주조사인구.}$$

(절차2) 청주시 범주별 실업자 총계 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

$$\text{청주}C1(l) = \text{청주시 1범주 실업자 총계,}$$

$$\text{청주}C2(l) = \text{청주시 2범주 실업자 총계,}$$

$$\text{청주}C3(l) = \text{청주시 3범주 실업자 총계,}$$

$$\text{청주}C4(l) = \text{청주시 4범주 실업자 총계.}$$

(절차3) 청주시 범주별 경제활동인구 총계 다시 카운트(l 번째 조사

구를 제외하고 카운트 함)

청주 $D1(l)$ = 청주시 1범주 경제활동인구 총계,

청주 $D2(l)$ = 청주시 2범주 경제활동인구 총계,

청주 $D3(l)$ = 청주시 3범주 경제활동인구 총계,

청주 $D4(l)$ = 청주시 4범주 경제활동인구 총계.

(절차4) 범주별 실업자 총계에 대한 직접추정값을 계산

○ 1범주 실업자 총계에 대한 직접추정값

$$\begin{aligned} {}_1\hat{Y}_{dir}(l) &= (\text{청주}A1/\text{청주}B1(l)) * \text{청주}C1(l) \\ &\quad + (\text{충주}A1/\text{충주}B1) * \text{충주}C1 \\ &\quad + (\text{제천}A1/\text{제천}B1) * \text{제천}C1 \end{aligned}$$

○ 2범주 직접추정값

$$\begin{aligned} {}_2\hat{Y}_{dir}(l) &= (\text{청주}A2/\text{청주}B2(l)) * \text{청주}C2(l) \\ &\quad + (\text{충주}A2/\text{충주}B2) * \text{충주}C2 \\ &\quad + (\text{제천}A2/\text{제천}B2) * \text{제천}C2 \end{aligned}$$

○ 3범주 직접추정값

$$\begin{aligned} {}_3\hat{Y}_{dir}(l) &= (\text{청주}A3/\text{청주}B3(l)) * \text{청주}C3(l) \\ &\quad + (\text{충주}A3/\text{충주}B3) * \text{충주}C3 \\ &\quad + (\text{제천}A3/\text{제천}B3) * \text{제천}C3 \end{aligned}$$

○ 4범주 직접추정값

$$\begin{aligned} {}_4\hat{Y}_{dir}(l) &= (\text{청주}A4/\text{청주}B4(l)) * \text{청주}C4(l) \\ &\quad + (\text{충주}A4/\text{충주}B4) * \text{충주}C4 \\ &\quad + (\text{제천}A4/\text{제천}B4) * \text{제천}C4 \end{aligned}$$

(절차5) l 번째 합성추정값 계산(= $\hat{Y}_1^s(l)$)

$$\begin{aligned} \hat{Y}_1^s(l) &= (\text{청주}A1 \times {}_1\hat{Y}_{dir}(l)) / (\text{청주}A1 + \text{충주}A1 + \text{제천}A1) \\ &\quad + (\text{청주}A2 \times {}_2\hat{Y}_{dir}(l)) / (\text{청주}A2 + \text{충주}A2 + \text{제천}A2) \end{aligned}$$

$$+(청주A3 \times_3 \hat{Y}_{\text{str}}(l)) / (청주A3 + 충주A3 + 제천A3)$$

$$+(청주A4 \times_4 \hat{Y}_{\text{str}}(l)) / (청주A4 + 충주A4 + 제천A4),$$

여기에서 $l = 1, 2, \dots, n_1$

(절차6) 청주시에서 생성된 n_1 개의 반복합성추정값들과 이들의 평균값을 저장

$$\hat{Y}_1^s(1), \hat{Y}_1^s(2), \dots, \hat{Y}_1^s(n_1),$$

$$\hat{Y}_1^s(\cdot) = (\hat{Y}_1^s(1) + \hat{Y}_1^s(2) + \dots + \hat{Y}_1^s(n_1)) / n_1$$

② 충주시에서 반복 합성추정량들을 생성

충주시에서 l 번째 표본조사구를 제외하고 나머지 표본조사구들을 이용하여 새로운 합성추정값을 다음 절차에 의해 반복적으로 생성함
($= \hat{Y}_2^s(l), l = 1, 2, \dots, n_2$)

(절차1) 충주시 범주별 상주조사인구 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

충주B1(l) = 충주시 1범주 상주조사인구,

충주B2(l) = 충주시 2범주 상주조사인구,

충주B3(l) = 충주시 3범주 상주조사인구,

충주B4(l) = 충주시 4범주 상주조사인구.

(절차2) 충주시 범주별 실업자 총계 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

충주C1(l) = 충주시 1범주 실업자 총계,

충주C2(l) = 충주시 2범주 실업자 총계,

충주C3(l) = 충주시 3범주 실업자 총계,

충주C4(l) = 충주시 4범주 실업자 총계.

(절차3) 충주시 범주별 경제활동인구 총계 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

충주 $D1(l)$ = 충주시 1범주 경제활동인구 총계,
 충주 $D2(l)$ = 충주시 2범주 경제활동인구 총계,
 충주 $D3(l)$ = 충주시 3범주 경제활동인구 총계,
 충주 $D4(l)$ = 충주시 4범주 경제활동인구 총계.

(절차4) 범주별 실업자 총계에 대한 직접추정값을 계산

o 1범주 실업자 총계에 대한 직접추정값

$$\begin{aligned}
 {}_1\hat{Y}_{dir}(l) &= (\text{충주}A1/\text{충주}B1(l)) * \text{충주}C1(l) \\
 &\quad + (\text{청주}A1/\text{청주}B1) * \text{청주}C1 \\
 &\quad + (\text{제천}A1/\text{제천}B1) * \text{제천}C1
 \end{aligned}$$

o 2범주 직접추정값

$$\begin{aligned}
 {}_2\hat{Y}_{dir}(l) &= (\text{충주}A2/\text{충주}B2(l)) * \text{충주}C2(l) \\
 &\quad + (\text{청주}A2/\text{청주}B2) * \text{청주}C2 \\
 &\quad + (\text{제천}A2/\text{제천}B2) * \text{제천}C2
 \end{aligned}$$

o 3범주 직접추정값

$$\begin{aligned}
 {}_3\hat{Y}_{dir}(l) &= (\text{충주}A3/\text{충주}B3(l)) * \text{충주}C3(l) \\
 &\quad + (\text{청주}A3/\text{청주}B3) * \text{청주}C3 \\
 &\quad + (\text{제천}A3/\text{제천}B3) * \text{제천}C3
 \end{aligned}$$

o 4범주 직접추정값

$$\begin{aligned}
 {}_4\hat{Y}_{dir}(l) &= (\text{충주}A4/\text{충주}B4(l)) * \text{충주}C4(l) \\
 &\quad + (\text{청주}A4/\text{청주}B4) * \text{청주}C4 \\
 &\quad + (\text{제천}A4/\text{제천}B4) * \text{제천}C4
 \end{aligned}$$

(절차5) l 번째 합성추정값 계산(= $\hat{Y}_2^S(l)$)

$$\begin{aligned}
 \hat{Y}_2^S(l) &= (\text{충주}A1 \times {}_1\hat{Y}_{dir}(l)) / (\text{청주}A1 + \text{충주}A1 + \text{제천}A1) \\
 &\quad + (\text{충주}A2 \times {}_2\hat{Y}_{dir}(l)) / (\text{청주}A2 + \text{충주}A2 + \text{제천}A2)
 \end{aligned}$$

$$+(\text{충주}A3 \times_3 \hat{Y}_{dr}(l))/(\text{충주}A3+\text{충주}A3+\text{제천}A3)$$

$$+(\text{충주}A4 \times_4 \hat{Y}_{dr}(l))/(\text{충주}A4+\text{충주}A4+\text{제천}A4),$$

여기에서 $l=1,2,\dots,n_2$

(절차6) 충주시에서 생성된 n_2 개의 반복합성추정값들과 이들의 평균값을 저장

$$\hat{Y}_2^s.(1), \hat{Y}_2^s.(2), \dots, \hat{Y}_2^s.(n_2),$$

$$\hat{Y}_2^s.(\cdot) = (\hat{Y}_2^s.(1) + \hat{Y}_2^s.(2) + \dots + \hat{Y}_2^s.(n_2)) / n_2$$

③ 제천시에서 반복 합성추정량들을 생성

제천시에서 l 번째 표본조사구를 제외하고 나머지 표본조사구들을 이용하여 새로운 합성추정값을 다음 절차에 의해 반복적으로 생성함
($= \hat{Y}_3^s.(l)$, $l=1,2,\dots,n_3$)

(절차1) 제천시 범주별 상주조사인구 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

$$\text{제천}B1(l) = \text{제천시 } 1\text{범주 상주조사인구},$$

$$\text{제천}B2(l) = \text{제천시 } 2\text{범주 상주조사인구},$$

$$\text{제천}B3(l) = \text{제천시 } 3\text{범주 상주조사인구},$$

$$\text{제천}B4(l) = \text{제천시 } 4\text{범주 상주조사인구}.$$

(절차2) 제주시 범주별 실업자 총계 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

$$\text{제천}C1(l) = \text{제천시 } 1\text{범주 실업자 총계},$$

$$\text{제천}C2(l) = \text{제천시 } 2\text{범주 실업자 총계},$$

$$\text{제천}C3(l) = \text{제천시 } 3\text{범주 실업자 총계},$$

$$\text{제천}C4(l) = \text{제천시 } 4\text{범주 실업자 총계}.$$

(절차3) 제천시 범주별 경제활동인구 총계 다시 카운트(l 번째 조사구를 제외하고 카운트 함)

제천 $D1(l)$ = 제천시 1범주 경제활동인구 총계,
 제천 $D2(l)$ = 제천시 2범주 경제활동인구 총계,
 제천 $D3(l)$ = 제천시 3범주 경제활동인구 총계,
 제천 $D4(l)$ = 제천시 4범주 경제활동인구 총계.

(절차4) 범주별 실업자 총계에 대한 직접추정값을 계산

○ 1범주 실업자 총계에 대한 직접추정값

$$\begin{aligned}
 {}_1\hat{Y}_{dir}(l) &= (\text{제천}A1/\text{제천}B1(l)) * \text{제천}C1(l) \\
 &\quad + (\text{청주}A1/\text{청주}B1) * \text{청주}C1 \\
 &\quad + (\text{충주}A1/\text{충주}B1) * \text{충주}C1
 \end{aligned}$$

○ 2범주 직접추정값

$$\begin{aligned}
 {}_2\hat{Y}_{dir}(l) &= (\text{제천}A2/\text{제천}B2(l)) * \text{제천}C2(l) \\
 &\quad + (\text{청주}A2/\text{청주}B2) * \text{청주}C2 \\
 &\quad + (\text{충주}A2/\text{충주}B2) * \text{충주}C2
 \end{aligned}$$

○ 3범주 직접추정값

$$\begin{aligned}
 {}_3\hat{Y}_{dir}(l) &= (\text{제천}A3/\text{제천}B3(l)) * \text{제천}C3(l) \\
 &\quad + (\text{청주}A3/\text{청주}B3) * \text{청주}C3 \\
 &\quad + (\text{충주}A3/\text{충주}B3) * \text{충주}C3
 \end{aligned}$$

○ 4범주 직접추정값

$$\begin{aligned}
 {}_4\hat{Y}_{dir}(l) &= (\text{제천}A4/\text{제천}B4(l)) * \text{제천}C4(l) \\
 &\quad + (\text{청주}A4/\text{청주}B4) * \text{청주}C4 \\
 &\quad + (\text{충주}A4/\text{충주}B4) * \text{제천}C4
 \end{aligned}$$

(절차5) l 번째 합성추정값 계산(= $\hat{Y}_3^s(l)$)

$$\begin{aligned}
 \hat{Y}_3^s(l) &= (\text{제천}A1 \times {}_1\hat{Y}_{dir}(l)) / (\text{청주}A1 + \text{충주}A1 + \text{제천}A1) \\
 &\quad + (\text{제천}A2 \times {}_2\hat{Y}_{dir}(l)) / (\text{청주}A2 + \text{충주}A2 + \text{제천}A2)
 \end{aligned}$$

$$+(\text{제천}A3 \times_3 \hat{Y}_{dr}(l)) / (\text{청주}A3 + \text{충주}A3 + \text{제천}A3)$$

$$+(\text{제천}A4 \times_4 \hat{Y}_{dr}(l)) / (\text{청주}A4 + \text{충주}A4 + \text{제천}A4),$$

여기에서 $l = 1, 2, \dots, n_3$

(절차6) 제천시에서 생성된 n_3 개의 반복합성추정값들과 이들의 평균값을 저장

$$\hat{Y}_{3.}^s(1), \hat{Y}_{3.}^s(2), \dots, \hat{Y}_{3.}^s(n_3),$$

$$\hat{Y}_{3.}^s(\cdot) = (\hat{Y}_{3.}^s(1) + \hat{Y}_{3.}^s(2) + \dots + \hat{Y}_{3.}^s(n_3)) / n_3$$

(Step4) 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)에서 반복 합성추정값들을 생성(Step3의 과정과 같음)

① 보은군에서 반복 합성추정값들을 생성

.....

⑧ 단양군에서 반복 합성추정값들을 생성

(Step5) 이상에서 반복 계산된 합성추정값들을 이용하여 충북내의 각 시군단위 소지역들에 대한 합성추정값들의 잭나이프 *Bias*, 잭나이프 분산, 잭나이프 *mse*, *CV* 값들을 계산하고 이 결과를 저장함.

충북 내의 시군 단위 소지역 i 에 대해서 다음을 계산하여 저장함.

$$\widehat{Bias}_j(\hat{Y}_{i.}^s) = (n_i - 1)[\hat{Y}_{i.}^s(\cdot) - \hat{Y}_{i.}^s], \quad i = 1, 2, \dots, 11$$

$$\widehat{Var}_j(\hat{Y}_{i.}^s) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} [\hat{Y}_{i.}^s(h) - \hat{Y}_{i.}^s(\cdot)]^2,$$

$$mse_j(\hat{Y}_{i.}^s) = \widehat{Var}_j(\hat{Y}_{i.}^s) + [\widehat{Bias}_j(\hat{Y}_{i.}^s)]^2,$$

$$CV(\hat{Y}_{i.}^s) = \frac{\sqrt{mse_j(\hat{Y}_{i.}^s)}}{\hat{Y}_{i.}^s} \times 100,$$

여기에서 $\hat{Y}_{i.}^s$ 는 (Step2)에 저장되어 있는 값들임.

(3) 복합추정법

(Step1) 각 시군 단위 소지역들에 대해서 최적 가중치 $\hat{w}_{i(opt)}$ 을 계산

$$\hat{w}_{i(opt)} = \frac{mse_J(\hat{Y}_{i.}^s)}{mse_J(\hat{Y}_{i.}^s) + \widehat{Var}(\hat{Y}_{i.})} \quad , \quad i=1,2,\dots,11$$

여기에서 $mse_J(\hat{Y}_{i.}^s)$ 는 (2)의 (Step5)에 저장되어 있고, $\widehat{Var}(\hat{Y}_{i.})$ 은 (1)의 (Step4)에 저장되어 있는 값들임.

(Step2) 각 시군 단위 소지역들에 대한 복합추정값 $\hat{Y}_{i.}^c$ 을 계산

$$\hat{Y}_{i.}^c = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s \quad , \quad i=1,2,\dots,11$$

여기에서 $\hat{Y}_{i.}$ 은 (1)의 (Step4)에 저장되어 있는 값들이고, $\hat{Y}_{i.}^s$ 는 (2)의 (Step2)에 저장되어 있는 값들임.

(Step3) 각 시군 단위 소지역 i 에 대해서 소지역 i 의 표본조사구 개수 ($= n_i$) 만큼 새로운 복합추정값들과 이들의 평균을 계산

소지역 i 에 대해서

$$\hat{Y}_{i.}^c (1) = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s (1), \quad i=1,2,\dots,11$$

$$\hat{Y}_{i.}^c (2) = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s (2),$$

...

$$\hat{Y}_{i.}^c (n_i) = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s (n_i),$$

$$\hat{Y}_{i.}^c (\cdot) = (\hat{Y}_{i.}^c (1) + \hat{Y}_{i.}^c (2) + \dots + \hat{Y}_{i.}^c (n_i)) / n_i$$

(Step4) 복합추정값들의 잭나이프 Bias, 잭나이프 분산, 잭나이프 mse, CV 값들을 계산하여 저장

소지역 i 에 대하여

$$\widehat{Bias}_J(\hat{Y}_{i.}^C) = (n_i - 1) [\hat{Y}_{i.}^C(\cdot) - \hat{Y}_{i.}^C], \quad i = 1, 2, \dots, 11$$

$$\widehat{Var}_J(\hat{Y}_{i.}^C) = \frac{n_i - 1}{n_i} \sum_{h=1}^n [\hat{Y}_{i.}^C(h) - \hat{Y}_{i.}^C]^2,$$

$$mse_J(\hat{Y}_{i.}^C) = \widehat{Var}_J(\hat{Y}_{i.}^C) + [\widehat{Bias}_J(\hat{Y}_{i.}^C)]^2,$$

$$CV(\hat{Y}_{i.}^C) = \frac{\sqrt{mse_J(\hat{Y}_{i.}^C)}}{\hat{Y}_{i.}^C} \times 100$$

여기에서 $\hat{Y}_{i.}^C(h)$ 와 $\hat{Y}_{i.}^C(\cdot)$ 은 위의 (Step3)에서 계산된 값들이고, $\hat{Y}_{i.}^C$ 은 위의 (Step2)에서 계산된 값들임.

(4) 복합추정값들의 총계 보정 및 추정결과 출력

(Step1) 시군구 단위 소지역 i 의 복합추정값 보정

대영역 내에서 분할된 3개의 시군구 그룹들 각각에 대해서 복합추정값들에 대한 총계 보정을 다음과 같이 실시한다.

① 시 그룹의 소지역들(청주, 충주, 제천)에 대한 총계 보정

$$\hat{Y}_{i.}^{C(adj)} = \frac{\hat{Y}_{i.}^C}{\sum_{i=1}^3 \hat{Y}_{i.}^C} \times \sum_{i=1}^3 \hat{Y}_{i.}^C, \quad i = 1, 2, 3$$

여기에서 $\sum_{i=1}^3 \hat{Y}_{i.}^C$ 은 시지역(청주, 충주, 제천시)의 직접추정값들에 대한 합을 나타낸다.

② 군 그룹의 소지역들(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양)에 대한 총계 보정

$$\hat{Y}_{i.}^{C(a\hat{d}_i)} = \frac{\hat{Y}_{i.}^C}{\sum_{i=1}^8 \hat{Y}_{i.}^C} \times \sum_{i=1}^8 \hat{Y}_{i.}, \quad i = 1, 2, \dots, 8$$

여기에서 $\sum_{i=1}^8 \hat{Y}_{i.}$ 은 군지역(보은, 옥천, 영동, 괴산, 진천, 음성, 청원, 단양군)의 직접추정값들에 대한 합을 나타낸다.

(Step2) 시군구 단위 추정결과 출력

구 분		직접추정법			합성추정법				복합추정법			
		$\hat{Y}_{i.}$	\sqrt{Var}	CV	$\hat{Y}_{i.}^s$	Bias	\sqrt{mse}	CV	$\hat{Y}_{i.}^c$	Bias	\sqrt{mse}	CV
시 지 역	청주시(1)											
	충주시(2)											
	제천시(3)											
군 지 역	보은군(4)											
	옥천군(5)											
	영동군(6)											
	괴산군(7)											
	음성군(8)											
	청원군(9)											
	진천군(10) 단양군(11)											
합												

5.7 추정결과

5.7.1 시군구 추정결과

2001년 10월의 경제활동인구조사 자료로부터 서울특별시 내의 구단위 소지역들의 실업자 총계 추정결과를 요약하면 다음 <표5.27>과 같이 주어진다.

<표5.27> 서울특별시의 구단위 실업자 총계 추정결과

지구	시군구	직접추정법			합성추정법			복합추정법		
		추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)
강남동	서초구	3,163	2,571	81.3	4,054	485	11.2	4,054	469	10.9
	강남구	4,756	2,516	53.0	5,709	606	9.9	5,694	573	9.5
	송파구	5,483	1,860	33.9	6,630	749	10.6	6,481	645	9.4
	강동구	7,942	3,577	45.0	4,951	790	14.9	5,114	753	13.9
강남서	양천구	1,570	1,200	76.5	5,781	169	2.6	5,721	166	2.6
	강서구	2,389	1,811	75.8	6,519	372	5.1	6,370	357	5.1
	구로구	11,832	2,796	23.6	5,367	520	8.7	5,567	503	8.2
	금천구	3,936	2,351	59.7	3,374	285	7.6	3,391	281	7.5
	영등포구	7,076	1,957	27.7	5,130	299	5.2	5,181	292	5.1
	동작구	3,936	699	17.8	5,282	181	3.1	5,194	170	3.0
	관악구	7,919	2,995	37.8	7,205	565	7.0	7,234	546	6.8
	성동구	9,420	1,881	20.0	5,840	298	4.6	5,921	290	4.4
강북동	광진구	9,489	2,934	30.9	6,837	522	6.9	6,909	506	6.6
	중랑구	5,506	1,687	30.6	7,539	255	3.0	7,503	249	3.0
	성북구	6,280	1,650	26.3	8,111	341	3.8	8,034	327	3.7
	강북구	3,959	2,245	56.7	6,160	371	5.4	6,109	361	5.3
	도봉구	4,687	1,526	32.6	6,063	225	3.3	6,041	221	3.3
	노원구	11,832	2,467	20.9	10,621	796	6.7	10,654	720	6.1
	종로구	1,570	1,176	75.0	1,732	74	4.0	1,749	74	4.0
강북서	중구	3,140	1,351	43.0	1,392	142	9.6	1,424	140	9.4
	용산구	1,570	1,002	63.9	2,371	244	9.8	2,346	230	9.4
	은평구	6,326	2,293	36.3	4,480	584	12.4	4,620	548	11.4
	서대문구	1,570	1,197	76.3	3,621	297	7.8	3,534	280	7.6
	마포구	5,506	2,349	42.7	3,754	429	10.8	3,841	415	10.3
	동대문구	1,570	1,035	66.0	3,902	314	7.6	3,738	287	7.4
계		132,427			132,427			132,427		

편의상 소지역 추정값들에 대한 CV값의 최대허용한계를 25%로 놓고 설명하기로 한다. 이 기준치는 현재 캐나다에서 대영역 내의 소지역 추정값들에 대해 적용하고 있는 CV값의 허용 기준치이다.

직접추정법을 적용하여 구단위 소지역들의 실업자 총계를 추정할 경우에는 “구로구”, “동작구”, “성동구”, “노원구”의 4개 지역만이 CV값의 목표허용한계를 만족하며, 나머지 21개 지역들은 CV값의 기준치인 25%를 벗어나는 매우 불안정한 추정결과를 나타낸다.

합성추정법과 복합추정법을 적용하여 구단위 소지역들의 실업자 총계를 추정할 경우에는 모든 지역들이 CV값의 목표허용한계를 만족하는 상당히 안정적인 추정결과를 나타낸다.

서울특별시의 구단위 소지역들에 대한 실업자 총계 추정결과, 직접추정값들은 대부분 신뢰할 수 없을 정도의 큰 추정오차를 동반한 반면 합성추정값과 복합추정값들은 매우 안정적인 추정결과를 나타냈으며, 효율 이득 면에서는 복합추정법이 합성추정법보다 탁월한 결과를 보였다.

나머지 대영역들도 서울특별시와 대동소이한 결과를 나타내며, 대영역별 시군구단위 소지역들의 추정결과를 정리하면 다음과 같다.

① 부산광역시

<표5.28> 부산광역시의 구/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
중구/기장군 중구 기장군	3,127	2,043	65.3	1,758	84	4.7	1,763 756 1,007	84	4.6
서구	1,042	1,150	110.3	2,082	68	3.2	2,081	68	3.2
영도구	1,536	1,257	81.8	2,507	129	5.0	2,500	128	5.0
동래구	4,663	1,060	22.7	4,029	198	4.8	4,052	191	4.6
남구	2,057	834	40.5	4,233	135	3.1	4,182	132	3.1
사하구	4,608	1,080	23.4	5,095	221	4.2	5,077	212	4.1
금정구	5,045	2,095	41.5	3,931	294	7.3	3,955	288	7.1
연제구	6,088	3,090	50.8	3,108	291	9.1	3,137	288	8.9
수영구	1,015	729	71.9	2,438	54	2.2	2,433	54	2.2
동구	2,057	1,118	54.3	1,594	70	4.1	1,602	69	4.0
부산진구	5,650	2,009	35.6	5,467	673	11.4	5,468	605	10.3
북구	4,635	1,197	25.8	3,732	239	6.0	3,769	230	5.7
해운대구	4,058	1,720	42.4	4,775	385	7.5	4,747	366	7.2
강서구	1,015	519	51.2	686	13	1.7	689	13	1.7
사상구	2,606	704	27.0	3,768	115	2.8	3,748	112	2.8
계	49,202			49,202			49,202		

② 인천광역시

<표5.29> 인천광역시의 구/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
중구	1,545	620	40.1	1,423	18	1.2	1,425	18	1.2
동구	1,156	393	34.0	1,551	16	0.9	1,552	16	0.9
남구	10,023	2,587	25.8	8,792	607	6.4	8,828	576	6.0
연수구	4,269	874	20.5	4,803	108	2.1	4,796	107	2.1
남동구	8,526	2,449	28.7	7,726	379	4.5	7,740	370	4.4
부평구	9,598	2,051	21.4	10,461	427	3.8	10,408	409	3.6
계양구	7,346	1,960	26.7	5,964	258	4.0	5,985	254	3.9
서구	5,412	1,429	26.4	6,402	181	2.6	6,388	179	2.6
강화군	389	314	80.8	1,139	9	0.7	1,140	9	0.7
계	48,264			48,264			48,264		

③ 대구광역시

<표5.30> 대구광역시의 구/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
중구	2,733	2,301	84.2	1,086	58	5.1	1,095	58	5.1
동구	3,632	943	26.0	4,100	169	4.0	4,110	164	3.9
서구	6,801	2,924	43.0	3,505	327	9.0	3,568	323	8.8
남구	2,719	911	33.5	2,429	76	3.0	2,448	75	3.0
북구	1,370	815	59.5	4,759	156	3.2	4,670	151	3.1
수성구	4,103	1,093	26.6	5,392	269	4.8	5,347	254	4.6
달서구	6,815	1,870	27.4	7,062	473	6.4	7,082	445	6.1
달성군	1,812	710	39.2	1,652	61	3.5	1,664	60	3.5
계	29,985			29,985			29,985		

④ 광주광역시

<표5.31> 광주광역시의 구단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
동구	0	0	0.0	1,233	0	0.0	1,145	0	0.0
서구	4,688	1,380	29.4	2,847	338	10.6	2,963	319	9.7
남구	2,483	1,576	63.5	2,368	260	9.8	2,398	253	9.5
북구	4,698	1,117	23.8	4,816	517	9.6	4,771	426	8.1
광산구	1,652	468	28.3	2,256	112	4.4	2,245	106	4.3
계	13,521			13,521			13,521		

⑤ 대전광역시

<표5.32> 대전광역시의 구단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
동구	3,277	1,530	46.7	3,516	274	7.7	3,517	265	7.5
중구	5,138	1,508	29.4	3,614	265	7.2	3,668	257	6.9
서구	5,821	1,440	24.7	6,572	491	7.4	6,503	440	6.7
유성구	1,441	1,085	75.3	2,406	117	4.8	2,401	116	4.8
대덕구	3,538	1,030	29.1	3,107	172	5.5	3,126	167	5.3
계	19,215			19,215			19,215		

⑥ 울산광역시

<표5.33> 울산광역시의 구/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
중구	3,674	1,473	40.1	2,922	448	13.8	2,999	410	12.5
남구	4,009	947	23.6	4,157	433	9.4	4,121	358	7.9
동구	1,988	653	32.8	2,299	154	6.0	2,307	146	5.8
울주군	1,720	726	42.2	2,013	324	14.5	1,964	270	12.5
계	11,391			11,391			11,391		

⑦ 경기도

<표5.34> 경기도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
평택시	4,366	2,887	66.1	2,391	702	20.9	2,517	662	19.4
용인시	1,363	1,484	108.9	3,262	522	11.4	3,116	465	11.0
파주시	1,789	1,010	56.5	1,528	555	25.8	1,521	426	20.6
이천시	989	749	75.8	1,288	234	12.9	1,280	213	12.3
안성시	1,007	782	77.7	1,044	45	3.1	1,079	45	3.1
광주시	3,732	1,749	46.9	1,305	202	20.6	1,474	200	19.6
포천군	711	634	89.3	1,986	304	20.3	1,951	247	18.3
가평/양평군 가평군 양평군	717	568	79.3	1,869	372	26.4	1,735 701 1,034	260	21.7
수원시	6,995	4,553	65.1	13,138	1,116	9.1	12,930	1,053	8.8
성남시	11,023	1,680	15.2	12,827	523	4.4	12,850	477	4.0
안양시	13,037	5,423	41.6	8,056	589	7.9	8,196	583	7.7
부천시	6,006	2,296	38.2	10,420	497	5.1	10,331	475	5.0
동두천시	5,017	2,374	47.3	948	11	1.3	956	11	1.3
고양시	8,056	2,491	30.9	9,498	469	5.3	9,551	453	5.1
과천/구리시 과천시 구리시	5,970	2,201	36.9	3,151	67	2.3	3,182 885 2,297	67	2.3
군포시	1,978	2,268	114.7	3,402	81	2.6	3,430	81	2.6
하남시	5,035	1,863	37.0	1,677	14	1.0	1,692	14	1.0
김포시	1,007	842	83.6	1,007	710	100.8	1,007	415	50.0
화성군	1,783	843	47.3	1,569	1,120	35.6	1,549	404	17.8
여주/연천군 여주군 연천군	1,072	622	58.0	1,286	547	21.2	1,307 874 433	308	16.1
의정부시	2,014	1,576	78.3	5,630	311	4.7	5,394	300	4.7
광명시	5,035	1,258	25.0	5,343	287	4.6	5,205	273	4.4
안산시	16,993	3,958	23.3	9,275	1,785	16.4	9,986	1,483	12.4
남양주시	2,856	3,809	133.4	5,333	332	5.3	5,225	329	5.3
오산시	2,967	1,700	57.3	1,789	49	2.3	1,761	49	2.3
시흥시	2,014	1,753	87.0	4,660	322	5.9	4,490	311	5.8
의왕시	2,014	798	39.7	1,864	26	1.2	1,833	26	1.2
양주군	2,150	1,279	59.5	2,150	3,577	72.8	2,150	405	16.5
계	117,696			117,696			117,696		

⑧ 강원도

<표5.35> 강원도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
춘천시	3,994	809	20.3	2,929	261	8.5	3,014	236	7.5
원주시	3,151	984	31.2	3,043	343	10.8	3,042	306	9.7
강릉시	2,587	1,137	43.9	2,592	267	9.9	2,587	253	9.4
동해시	1,586	1,002	63.2	1,057	79	7.2	1,060	78	7.1
태백시	0	0	0.0	603	0	0.0	531	0	0.0
속초시	401	324	80.8	928	29	3.0	923	29	3.0
삼척시	243	226	93.0	810	23	2.7	805	23	2.8
홍천군	0	0	0.0	225	0	0.0	213	0	0.0
횡성/영월 /고성군	705	318	45.1	404	223	39.5	443	149	24.4
횡성군							153		
영월군							174		
고성군							116		
평창/정선 /양양군	0	0	0.0	411	0	0.0	390	0	0.0
평창군							144		
정선군							152		
양양군							94		
철원/화천군	0	0	0.0	233	0	0.0	220	0	0.0
철원군							149		
화천군							71		
양구/인제군	728	444	61.0	160	47	21.1	167	47	20.3
양구군							69		
인제군							98		
계	13,395			13,395			13,395		

⑨ 충청북도

<표5.36> 충청북도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
청주시	4,740	1,305	27.5	4,475	1,090	23.0	4,502	642	13.5
충주시	1,035	502	48.5	1,543	152	9.3	1,505	139	8.8
제천시	1,247	640	51.3	1,004	109	10.3	1,015	106	9.9
청원군	553	419	75.8	675	87	10.2	675	83	10.0
보은/영동군	831	378	45.4	521	73	11.2	534	71	10.7
보은군							226		
영동군							308		
옥천/진천군	831	640	76.9	616	136	17.6	627	130	16.8
옥천군							309		
진천군							318		
괴산군	0	0	0.0	339	0	0.0	313	0	0.0
음성군	274	315	115.1	436	51	9.3	437	49	9.1
단양군	274	312	113.9	177	11	5.1	179	11	5.1
계	9,785			9,785			9,785		

⑩ 충청남도

<표5.37> 충청남도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
천안시	6,096	1,233	20.2	4,903	429	7.7	4,968	383	6.8
공주시	1,138	687	60.3	1,478	110	6.6	1,470	108	6.5
보령시	694	331	47.7	1,168	44	3.3	1,161	43	3.3
아산시	2,452	991	40.4	2,124	159	6.6	2,130	155	6.4
서산시	2,640	1,185	44.9	1,537	144	8.3	1,552	142	8.1
논산시	0	0	0.0	1,810	0	0.0	1,740	0	0.0
금산/서천 /청양군	978	581	59.4	931	136	12.7	931	129	12.1
금산군							334		
서천군							386		
청양군							211		
연기/홍성군	1,486	1,520	102.3	1,191	556	40.6	1,208	490	35.5
연기군							563		
홍성군							645		
부여군	471	384	81.7	462	45	8.4	463	44	8.3
예산/태안군	978	567	58.0	1,066	189	15.5	1,049	170	14.2
예산군							626		
태안군							423		
당진군	471	519	110.3	733	56	6.7	732	56	6.6
계	17,404			17,404			17,404		

⑪ 전라북도

<표5.38> 전라북도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
전주시	8,303	1,857	22.4	6,634	806	12.2	6,842	678	9.9
군산시	3,019	1,435	47.5	2,986	234	7.9	2,959	228	7.7
익산시	2,642	952	36.0	3,463	388	11.3	3,317	333	10.0
정읍시	528	639	120.9	1,330	60	4.5	1,311	59	4.5
남원시	1,484	517	34.9	996	61	6.2	993	60	6.0
김제시	528	611	115.6	1,094	34	3.1	1,082	34	3.1
완주군	0	0	0.0	525	0	0.0	454	0	0.0
진안/장수 /임실군	1,272	314	24.7	438	126	23.7	497	109	17.1
진안군							162		
장수군							151		
임실군							184		
무주/순창 /고창군	0	0	0.0	616	0	0.0	533	0	0.0
무주군							117		
순창군							131		
고창군							285		
부안군	645	288	44.6	337	356	87.3	433	25	25.5
계	18,421			18,421			18,421		

⑫ 전라남도

<표5.39> 전라남도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
목포시	7,636	1,153	15.1	4,196	409	10.3	4,541	364	8.3
여수시	5,051	1,516	30.0	4,643	373	8.5	4,592	352	8.0
순천시	2,729	708	25.9	4,126	231	5.9	3,929	209	5.5
나주시	444	351	79.0	1,478	30	2.2	1,442	30	2.2
광양시	494	461	93.2	1,912	62	3.5	1,850	61	3.4
담양군	494	470	95.0	339	26	7.0	339	26	7.0
곡성/구례 /보성군	494	512	103.5	708	55	7.0	705	54	7.0
곡성군							209		
구례군							177		
보성군							319		
고흥군	1,382	833	60.2	502	89	16.2	511	88	15.8
화순/장흥군	1,876	1,227	65.4	697	309	40.4	758	290	35.0
화순군							459		
장흥군							299		
강진군	494	439	88.9	261	20	7.1	261	20	7.1
해남군	494	357	72.1	513	38	6.8	512	38	6.7
영암군	0	0	0.0	352	0	0.0	341	0	0.0
무안군	0	0	0.0	378	0	0.0	367	0	0.0
함평군	0	0	0.0	241	0	0.0	234	0	0.0
영광/완도 /진도군	0	0	0.0	947	0	0.0	919	0	0.0
영광군							365		
완도군							340		
진도군							214		
장성군	0	0	0.0	298	0	0.0	289	0	0.0
계	21,588			21,588			21,588		

⑬ 경상북도

<표5.40> 경상북도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
포항시	5,790	1,271	22.0	6,009	429	6.3	5,929	385	5.8
경주시	2,393	1,712	71.6	3,102	183	5.3	3,095	181	5.2
김천시	2,062	1,858	90.1	1,744	125	6.4	1,746	124	6.3
안동시	516	609	118.3	2,132	49	2.1	2,124	49	2.1
구미시	8,457	2,336	27.6	4,141	394	8.5	4,240	383	8.0
영주시	932	414	44.4	1,447	41	2.5	1,443	40	2.5
영천시	516	522	101.2	1,220	27	2.0	1,219	27	2.0
상주시	956	679	71.0	1,242	25	1.8	1,243	24	1.8
문경시	977	909	93.1	1,074	49	4.1	1,074	49	4.1
경산시	2,339	1,592	68.1	2,826	164	5.1	2,822	162	5.1
군위/영양 /성주군	0	0	0.0	238	0	0.0	191	0	0.0
군위군							61		
영양군							40		
성주군							90		
의성/청송군	0	0	0.0	228	0	0.0	182	0	0.0
의성군							125		
청송군							57		
영덕/청도군	471	542	115.1	224	171	57.0	263	156	49.2
영덕군							129		
청도군							134		
고령/봉화군	0	0	0.0	176	0	0.0	141	0	0.0
고령군							67		
봉화군							74		
칠곡/예천 /울진군	911	601	66.0	517	253	36.4	605	215	29.5
칠곡군							280		
예천군							152		
울진군							173		
계	26,320			26,320			26,320		

⑭ 경상남도

<표5.41> 경상남도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
창원시	5,819	2,453	42.2	7,543	369	4.9	7,542	360	4.8
마산시	7,208	2,209	30.6	6,987	451	6.5	7,032	433	6.2
진주시	1,426	592	41.5	5,291	134	2.5	5,128	127	2.5
진해시	2,815	725	25.7	1,974	45	2.3	1,987	45	2.3
통영시	3,573	1,066	29.8	1,927	43	2.2	1,939	43	2.2
사천시	360	371	103.1	1,726	24	1.4	1,729	24	1.4
김해시	8,660	2,009	23.2	4,924	273	5.6	5,017	268	5.4
밀양시	3,547	3,355	94.6	1,808	150	8.3	1,820	149	8.3
거제시	2,461	781	31.7	2,575	78	3.1	2,587	77	3.0
양산시	1,793	747	41.7	2,907	141	4.9	2,883	136	4.8
의령/합천군	0	0	0.0	334	0	0.0	292	0	0.0
의령군							105		
합천군							187		
함안/하동/거창군	1,441	653	45.3	783	573	35.7	862	324	21.1
함안군							292		
하동군							266		
거창군							304		
창녕/고성	0	0	0.0	509	0	0.0	445	0	0.0
창녕군							237		
고성군							208		
남해/산청/함양군	720	511	70.9	535	301	27.5	562	224	22.4
남해군							227		
산청군							157		
함양군							178		
계	39,823			39,823			39,823		

⑮ 제주도

<표5.42> 제주도의 시/군단위 실업자 총계 추정결과

시군구	직접추정법			합성추정법			복합추정법		
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
제주시	1,931	797	41.3	1,698	584	31.7	1,719	380	20.3
서귀포시	243	194	79.9	476	56	10.8	455	52	10.4
북제주군	0	0	0.0	0	0	0.0	0	0	0.0
남제주군	0	0	0.0	0	0	0.0	0	0	0.0
계	2,174			2,174			2,174		

5.7.2 효율비교

2002년 10월 서울특별시의 구단위 소지역들의 추정결과를 기반으로 직접추정값 대비 합성 또는 복합추정값들의 상대효율이득을 <표5.43>에 요약하였다. 직접추정값 대비 합성 또는 복합추정값들의 상대효율이득은 다음 식으로 주어진다.

$$\text{상대효율이득} = \frac{\text{직접추정값의 CV값} - \text{합성/복합추정값의 CV값}}{\text{직접추정값의 CV값}} * 100(\%)$$

<표5.43> 직접추정 대비 합성/복합추정의 상대효율이득

지구	시군구	직접추정법		합성추정법			복합추정법		
		추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
강남동	서초구	3,163	81.3	4,054	11.2	86.2	4,054	10.9	86.6
	강남구	4,756	53.0	5,709	9.9	81.3	5,694	9.5	82.1
	송파구	5,483	33.9	6,630	10.6	68.7	6,481	9.4	72.3
	강동구	7,942	45.0	4,951	14.9	66.9	5,114	13.9	69.1
강남서	양천구	1,570	76.5	5,781	2.6	96.6	5,721	2.6	96.6
	강서구	2,389	75.8	6,519	5.1	93.3	6,370	5.1	93.3
	구로구	11,832	23.6	5,367	8.7	63.1	5,567	8.2	65.3
	금천구	3,936	59.7	3,374	7.6	87.3	3,391	7.5	87.4
	영등포구	7,076	27.7	5,130	5.2	81.2	5,181	5.1	81.6
	동작구	3,936	17.8	5,282	3.1	82.6	5,194	3.0	83.1
강북동	관악구	7,919	37.8	7,205	7.0	81.5	7,234	6.8	82.0
	성동구	9,420	20.0	5,840	4.6	77.0	5,921	4.4	78.0
	광진구	9,489	30.9	6,837	6.9	77.7	6,909	6.6	78.6
	중랑구	5,506	30.6	7,539	3.0	90.2	7,503	3.0	90.2
	성북구	6,280	26.3	8,111	3.8	85.6	8,034	3.7	85.9
	강북구	3,959	56.7	6,160	5.4	90.5	6,109	5.3	90.7
	도봉구	4,687	32.6	6,063	3.3	89.9	6,041	3.3	89.9
	노원구	11,832	20.9	10,621	6.7	67.9	10,654	6.1	70.8
강북서	종로구	1,570	75.0	1,732	4.0	94.7	1,749	4.0	94.7
	중구	3,140	43.0	1,392	9.6	77.7	1,424	9.4	78.1
	용산구	1,570	63.9	2,371	9.8	84.7	2,346	9.4	85.3
	은평구	6,326	36.3	4,480	12.4	65.8	4,620	11.4	68.6
	서대문구	1,570	76.3	3,621	7.8	89.8	3,534	7.6	90.0
	마포구	5,506	42.7	3,754	10.8	74.7	3,841	10.3	75.9
	동대문구	1,570	66.0	3,902	7.6	88.5	3,738	7.4	88.8
계	132,427		132,427		평균 81.7	132,427		평균 82.6	

직접추정 대비 합성 및 복합추정의 상대효율이득을 살펴보자. 합성추정법을 적용하여 실업자 총계를 추정할 경우, 직접추정 결과에 비해 최소 63.1%(구로구)에서 최대 96.6%(양천구)의 상대효율이득이 발생하였고, 전체 지역에 대해서는 평균적으로 81.0%의 상대효율이득이 발생하였다. 복합추정법을 적용할 경우에는 직접추정법에 비해 최소 49.8%(구로구)에서 최대 95.8%(양천구)의 상대효율이득이 발생하였고, 전체 지역에 대해서는 평균적으로 82.8%의 상대효율이득이 발생하였다. 서울특별시의 경우, 합성추정 및 복합추정 공히 신뢰할 만한 추정결과를 보여주며, 효율이득 면에서는 복합추정법에 의한 추정결과가 탁월하게 나타났다.

서울특별시 이외의 나머지 15개 대영역들도 서울특별시와 마찬가지로 유사한 추정결과를 나타내며 대영역별 상대효율이득을 정리하면 다음과 같다.

① 부산광역시

<표5.44> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
중구/기장군 중구 기장군	3,127	65.3	1,758	4.7	92.8	1,763 756 1,007	4.6	93.0
서구	1,042	110.3	2,082	3.2	97.1	2,081	3.2	97.1
영도구	1,536	81.8	2,507	5.0	93.9	2,500	5.0	93.9
동래구	4,663	22.7	4,029	4.8	78.9	4,052	4.6	79.7
남구	2,057	40.5	4,233	3.1	92.3	4,182	3.1	92.3
사하구	4,608	23.4	5,095	4.2	82.1	5,077	4.1	82.5
금정구	5,045	41.5	3,931	7.3	82.4	3,955	7.1	82.9
연제구	6,088	50.8	3,108	9.1	82.1	3,137	8.9	82.5
수영구	1,015	71.9	2,438	2.2	96.9	2,433	2.2	96.9
동구	2,057	54.3	1,594	4.1	92.4	1,602	4.0	92.6
부산진구	5,650	35.6	5,467	11.4	68.0	5,468	10.3	71.1
북구	4,635	25.8	3,732	6.0	76.7	3,769	5.7	77.9
해운대구	4,058	42.4	4,775	7.5	82.3	4,747	7.2	83.0
강서구	1,015	51.2	686	1.7	96.7	689	1.7	96.7
사상구	2,606	27.0	3,768	2.8	89.6	3,748	2.8	89.6
계	49,202		49,202		평균 86.9	49,202		평균 87.4

② 인천광역시

<표5.45> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
중구	1,545	40.1	1,423	1.2	97.0	1,425	1.2	97.0
동구	1,156	34.0	1,551	0.9	97.4	1,552	0.9	97.4
남구	10,023	25.8	8,792	6.4	75.2	8,828	6.0	76.7
연수구	4,269	20.5	4,803	2.1	89.8	4,796	2.1	89.8
남동구	8,526	28.7	7,726	4.5	84.3	7,740	4.4	84.7
부평구	9,598	21.4	10,461	3.8	82.2	10,408	3.6	83.2
계양구	7,346	26.7	5,964	4.0	85.0	5,985	3.9	85.4
서구	5,412	26.4	6,402	2.6	90.2	6,388	2.6	90.2
강화군	389	80.8	1,139	0.7	99.1	1,140	0.7	99.1
계	48,264		48,264		평균 88.9	48,264		평균 89.3

③ 대구광역시

<표5.46> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
중구	2,733	84.2	1,086	5.1	93.9	1,095	5.1	93.9
동구	3,632	26.0	4,100	4.0	84.6	4,110	3.9	85.0
서구	6,801	43.0	3,505	9.0	79.1	3,568	8.8	79.5
남구	2,719	33.5	2,429	3.0	91.0	2,448	3.0	91.0
북구	1,370	59.5	4,759	3.2	94.6	4,670	3.1	94.8
수성구	4,103	26.6	5,392	4.8	82.0	5,347	4.6	82.7
달서구	6,815	27.4	7,062	6.4	76.6	7,082	6.1	77.7
달성군	1,812	39.2	1,652	3.5	91.1	1,664	3.5	91.1
계	29,985		29,985		평균 86.6	29,985		평균 87.0

④ 광주광역시

<표5.47> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
동구	0	0.0	1,233	0.0	-	1,145	0.0	-
서구	4,688	29.4	2,847	10.6	63.9	2,963	9.7	67.0
남구	2,483	63.5	2,368	9.8	84.6	2,398	9.5	85.0
북구	4,698	23.8	4,816	9.6	59.7	4,771	8.1	66.0
광산구	1,652	28.3	2,256	4.4	84.5	2,245	4.3	84.8
계	13,521		13,521		평균 73.2	13,521		평균 75.7

⑤ 대전광역시

<표5.48> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
동구	3,277	46.7	3,516	7.7	83.5	3,517	7.5	83.9
중구	5,138	29.4	3,614	7.2	75.5	3,668	6.9	76.5
서구	5,821	24.7	6,572	7.4	70.0	6,503	6.7	72.9
유성구	1,441	75.3	2,406	4.8	93.6	2,401	4.8	93.6
대덕구	3,538	29.1	3,107	5.5	81.1	3,126	5.3	81.8
계	19,215		19,215		평균 80.7	19,215		평균 81.7

⑥ 울산광역시

<표5.49> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
중구	3,674	40.1	2,922	13.8	65.6	2,999	12.5	68.8
남구	4,009	23.6	4,157	9.4	60.2	4,121	7.9	66.5
동구	1,988	32.8	2,299	6.0	81.7	2,307	5.8	82.3
울주군	1,720	42.2	2,013	14.5	65.6	1,964	12.5	70.4
계	11,391		11,391		평균 68.3	11,391		평균 72.0

⑦ 경기도

<표5.50> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
평택시	4,366	66.1	2,391	20.9	68.4	2,517	19.4	70.7
용인시	1,363	108.9	3,262	11.4	89.5	3,116	11.0	89.9
파주시	1,789	56.5	1,528	25.8	54.3	1,521	20.6	63.5
이천시	989	75.8	1,288	12.9	83.0	1,280	12.3	83.8
안성시	1,007	77.7	1,044	3.1	96.0	1,079	3.1	96.0
광주시	3,732	46.9	1,305	20.6	56.1	1,474	19.6	58.2
포천군	711	89.3	1,986	20.3	77.3	1,951	18.3	79.5
가평/양평군 가평군 양평군	717	79.3	1,869	26.4	66.7	1,735 701 1,034	21.7	72.6
수원시	6,995	65.1	13,138	9.1	86.0	12,930	8.8	86.5
성남시	11,023	15.2	12,827	4.4	71.1	12,850	4.0	73.7
안양시	13,037	41.6	8,056	7.9	81.0	8,196	7.7	81.5
부천시	6,006	38.2	10,420	5.1	86.6	10,331	5.0	86.9
동두천시	5,017	47.3	948	1.3	97.3	956	1.3	97.3
고양시	8,056	30.9	9,498	5.3	82.8	9,551	5.1	83.5
과천/구리시 과천시 구리시	5,970	36.9	3,151	2.3	93.8	3,182 885 2,297	2.3	93.8
군포시	1,978	114.7	3,402	2.6	97.7	3,430	2.6	97.7
하남시	5,035	37.0	1,677	1.0	97.3	1,692	1.0	97.3
김포시	1,007	83.6	1,007	100.8	-20.6	1,007	50.0	40.2
화성군	1,783	47.3	1,569	35.6	24.7	1,549	17.8	62.4
여주/연천군 여주군 연천군	1,072	58.0	1,286	21.2	63.4	1,307 874 433	16.1	72.2
의정부시	2,014	78.3	5,630	4.7	94.0	5,394	4.7	94.0
광명시	5,035	25.0	5,343	4.6	81.6	5,205	4.4	82.4
안산시	16,993	23.3	9,275	16.4	29.6	9,986	12.4	46.8
남양주시	2,856	133.4	5,333	5.3	96.0	5,225	5.3	96.0
오산시	2,967	57.3	1,789	2.3	96.0	1,761	2.3	96.0
시흥시	2,014	87.0	4,660	5.9	93.2	4,490	5.8	93.3
의왕시	2,014	39.7	1,864	1.2	97.0	1,833	1.2	97.0
양주군	2,150	59.5	2,150	72.8	-22.4	2,150	16.5	72.3
계	117,696		117,696		평균 72.1	117,696		평균 80.9

⑧ 강원도

<표5.51> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
춘천시	3,994	20.3	2,929	8.5	58.1	3,014	7.5	63.1
원주시	3,151	31.2	3,043	10.8	65.4	3,042	9.7	68.9
강릉시	2,587	43.9	2,592	9.9	77.4	2,587	9.4	78.6
동해시	1,586	63.2	1,057	7.2	88.6	1,060	7.1	88.8
태백시	0	0.0	603	0.0	-	531	0.0	-
속초시	401	80.8	928	3.0	96.3	923	3.0	96.3
삼척시	243	93.0	810	2.7	97.1	805	2.7	97.1
홍천군	0	0.0	225	0.0	-	213	0.0	-
횡성/영월 /고성군	705	45.1	404	39.5	12.4	443	24.4	45.9
횡성군						153		
영월군						174		
고성군						116		
평창/정선 /양양군	0	0.0	411	0.0	-	390	0.0	-
평창군						144		
정선군						152		
양양군						94		
철원/화천군	0	0.0	233	0.0	-	220	0.0	-
철원군						149		
화천군						71		
양구/인제군	728	61.0	160	21.1	65.4	167	20.3	66.7
양구군						69		
인제군						98		
계	13,395		13,395		평균70.1	13,395		평균75.7

⑨ 충청북도

<표5.52> 직접추정 대비 합성/복합추정의 상대효율이익

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이익(%)	추정값	CV (%)	효율 이익(%)
청주시	4,740	27.5	4,475	23.0	16.4	4,502	13.5	50.9
충주시	1,035	48.5	1,543	9.3	80.8	1,505	8.8	81.9
재천시	1,247	51.3	1,004	10.3	79.9	1,015	9.9	80.7
청원군	553	75.8	675	10.2	86.5	675	10.0	86.8
보은/영동군	831	45.4	521	11.2	75.3	534	10.7	76.4
보은군						226		
영동군						308		
옥천/진천군	831	76.9	616	17.6	77.1	627	16.8	78.2
옥천군						309		
진천군						318		
괴산군	0	0.0	339	0.0	-	313	0.0	-
음성군	274	115.1	436	9.3	91.9	437	9.1	92.1
단양군	274	113.9	177	5.1	95.5	179	5.1	95.5
계	9,785		9,785		평균75.4	9,785		평균80.3

⑩ 충청남도

<표5.53> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
천안시	6,096	20.2	4,903	7.7	61.9	4,968	6.8	66.3
공주시	1,138	60.3	1,478	6.6	89.1	1,470	6.5	89.2
보령시	694	47.7	1,168	3.3	93.1	1,161	3.3	93.1
아산시	2,452	40.4	2,124	6.6	83.7	2,130	6.4	84.2
서산시	2,640	44.9	1,537	8.3	81.5	1,552	8.1	82.0
논산시	0	0.0	1,810	0.0	-	1,740	0.0	-
금산/서천 /청양군	978	59.4	931	12.7	78.6	931	12.1	79.6
금산군						334		
서천군						386		
청양군						211		
연기/홍성군 연기군 홍성군	1,486	102.3	1,191	40.6	60.3	1,208	35.5	65.3
연기군						563		
홍성군						645		
부여군	471	81.7	462	8.4	89.7	463	8.3	89.8
예산/태안군 예산군 태안군	978	58.0	1,066	15.5	73.3	1,049	14.2	75.5
예산군						626		
태안군						423		
당진군	471	110.3	733	6.7	93.9	732	6.6	94.0
계	17,404		17,404		평균80.5	17,404		평균81.9

㉠ 전라북도

<표5.54> 직접추정 대비 합성/복합추정의 상대효율이익

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이익(%)	추정값	CV (%)	효율 이익(%)
전주시	8,303	22.4	6,634	12.2	45.5	6,842	9.9	55.8
군산시	3,019	47.5	2,986	7.9	83.4	2,959	7.7	83.8
익산시	2,642	36.0	3,463	11.3	68.6	3,317	10.0	72.2
정읍시	528	120.9	1,330	4.5	96.3	1,311	4.5	96.3
남원시	1,484	34.9	996	6.2	82.2	993	6.0	82.8
김제시	528	115.6	1,094	3.1	97.3	1,082	3.1	97.3
완주군	0	0.0	525	0.0	-	454	0.0	-
진안/장수 /임실군	1,272	24.7	438	23.7	4.0	497	17.1	30.8
진안군						162		
장수군						151		
임실군						184		
무주/순창 /고창군	0	0.0	616	0.0	-	533	0.0	-
무주군						117		
순창군						131		
고창군						285		
부안군	645	44.6	337	87.3	-95.7	433	25.5	42.8
계	18,421		18,421		평균47.7	18,421		평균70.4

⑫ 전라남도

<표5.55> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
목포시	7,636	15.1	4,196	10.3	31.8	4,541	8.3	45.0
여수시	5,051	30.0	4,643	8.5	71.7	4,592	8.0	73.3
순천시	2,729	25.9	4,126	5.9	77.2	3,929	5.5	78.8
나주시	444	79.0	1,478	2.2	97.2	1,442	2.2	97.2
광양시	494	93.2	1,912	3.5	96.2	1,850	3.4	96.4
담양군	494	95.0	339	7.0	92.6	339	7.0	92.6
곡성/구례 /보성군	494	103.5	708	7.0	93.2	705	7.0	93.2
곡성군						209		
구례군						177		
보성군						319		
고흥군	1,382	60.2	502	16.2	73.1	511	15.8	73.8
화순/장흥군	1,876	65.4	697	40.4	38.2	758	35.0	46.5
화순군						459		
장흥군						299		
강진군	494	88.9	261	7.1	92.0	261	7.1	92.0
해남군	494	72.1	513	6.8	90.6	512	6.7	90.7
영암군	0	0.0	352	0.0	-	341	0.0	-
무안군	0	0.0	378	0.0	-	367	0.0	-
함평군	0	0.0	241	0.0	-	234	0.0	-
영광/완도 /진도군	0	0.0	947	0.0	-	919	0.0	-
영광군						365		
완도군						340		
진도군						214		
장성군	0	0.0	298	0.0	-	289	0.0	-
계	21,588		21,588		평균77.6	21,588		평균80.0

⑬ 경상북도

<표5.56> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
포항시	5,790	22.0	6,009	6.3	71.4	5,929	5.8	73.6
경주시	2,393	71.6	3,102	5.3	92.6	3,095	5.2	92.7
김천시	2,062	90.1	1,744	6.4	92.9	1,746	6.3	93.0
안동시	516	118.3	2,132	2.1	98.2	2,124	2.1	98.2
구미시	8,457	27.6	4,141	8.5	69.2	4,240	8.0	71.0
영주시	932	44.4	1,447	2.5	94.4	1,443	2.5	94.4
영천시	516	101.2	1,220	2.0	98.0	1,219	2.0	98.0
상주시	956	71.0	1,242	1.8	97.5	1,243	1.8	97.5
문경시	977	93.1	1,074	4.1	95.6	1,074	4.1	95.6
경산시	2,339	68.1	2,826	5.1	92.5	2,822	5.1	92.5
군위/영양 /성주군	0	0.0	238	0.0	-	191	0.0	-
군위군						61		
영양군						40		
성주군						90		
의성/청송군	0	0.0	228	0.0	-	182	0.0	-
의성군						125		
청송군						57		
영덕/청도군	471	115.1	224	57.0	50.5	263	49.2	57.3
영덕군						129		
청도군						134		
고령/봉화군	0	0.0	176	0.0	-	141	0.0	-
고령군						67		
봉화군						74		
칠곡/예천 /울진군	911	66.0	517	36.4	44.8	605	29.5	55.3
칠곡군						280		
예천군						152		
울진군						173		
계	26,320		26,320		평균83.1	26,320		평균84.9

⑭ 경상남도

<표5.57> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
창원시	5,819	42.2	7,543	4.9	88.4	7,542	4.8	88.6
마산시	7,208	30.6	6,987	6.5	78.8	7,032	6.2	79.7
진주시	1,426	41.5	5,291	2.5	94.0	5,128	2.5	94.0
진해시	2,815	25.7	1,974	2.3	91.1	1,987	2.3	91.1
통영시	3,573	29.8	1,927	2.2	92.6	1,939	2.2	92.6
사천시	360	103.1	1,726	1.4	98.6	1,729	1.4	98.6
김해시	8,660	23.2	4,924	5.6	75.9	5,017	5.4	76.7
밀양시	3,547	94.6	1,808	8.3	91.2	1,820	8.3	91.2
거제시	2,461	31.7	2,575	3.1	90.2	2,587	3.0	90.5
양산시	1,793	41.7	2,907	4.9	88.2	2,883	4.8	88.5
의령/합천군	0	0.0	334	0.0	-	292	0.0	-
의령군						105		
합천군						187		
합안/하동/거창군	1,441	45.3	783	35.7	21.2	862	21.1	53.4
합안군						292		
하동군						266		
거창군						304		
창녕/고성군	0	0.0	509	0.0	-	445	0.0	-
창녕군						237		
고성군						208		
남해/산청/함양군	720	70.9	535	27.5	61.2	562	22.4	68.4
남해군						227		
산청군						157		
함양군						178		
계	39,823		39,823		평균81.0	39,823		평균84.4

⑮ 제주도

<표5.58> 직접추정 대비 합성/복합추정의 상대효율이득

시군구	직접추정법		합성추정법			복합추정법		
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)
제주시	1,931	41.3	1,698	31.7	23.2	1,719	20.3	50.8
서귀포시	243	79.9	476	10.8	86.5	455	10.4	87.0
북제주군	0	0.0	0	0.0	-	0	0.0	-
남제주군	0	0.0	0	0.0	-	0	0.0	-
계	2,174		2,174		평균54.9	2,174		평균68.9

제6장 시·도단위의 세부영역 통계 작성

6.1 개요

시군구단위 소지역들에 대한 실업자, 취업자, 경제활동인구 및 비경제활동인구의 월별 통계자료는 정부의 실업자 구제정책 마련과 지역경제의 균형발전 등에 매우 실용적인 참고자료로 활용될 수 있다. 이와 더불어 대영역에 대한 “성별-연령대별” 또는 “성별-교육정도별” 범주들에 대한 월별 경제활동 특성을 파악하여 실업자 구제정책 입안을 위한 보조정보로 활용할 수 있다면 정부 정책은 보다 구체화될 수 있다.

이번 장에서는 제 5장에서 소개한 소지역추정법을 이용하여 특·광역시 및 도지역들의 세부영역별 실업자 총계를 추정하고자 한다. 대영역별 세부영역은 “성별(남, 여)-연령대별(15-29세, 30-39세, 40-49세, 50세 이상)”의 8개 범주와 “성별(남, 여)-교육정도별(초등졸 이하, 중졸, 고졸, 대졸이상)”의 8개 범주로써 총 16개 범주이다.

합성추정법을 이용하여 범주별 실업자 총계를 추정하기 위한 “Borrow Strength” 산정 절차는 다음과 같다. 우선 2002년 「인구주택 총 조사」 자료를 이용하여 16개 특·광역시 및 도단위 대영역들을 선정된 연구변수들에 의해 경제활동 특성이 유사한 동질적인 부차그룹들로 분할한다. 이러한 부차그룹들은 동질적인 경제활동 특성을 갖는 대영역들로 구성된다. 다음으로 경제활동인구조사 자료로부터 각 부차그룹에 대한 “성별-연령대별” 8개 범주와 “성별-교육정도별” 8개 범주에 대한 실업자 총계를 추정한다. 부차그룹 내의 “성별-연령대별” 및 “성별-교육정도별” 세부범주들에 대한 실업자 총계 추정값들을 “Borrow Strength”로 활용한다.

“Borrow Strength” 적용 기준은 다음과 같다. 각 대영역에 대한 “성별-연령대별” 8개 범주의 실업자 총계 추정 시 해당 대영역을 포함하고 있는 부차그룹의 “성별-교육정도별” 8개 범주의 실업자 총계 추정값들을 “Borrow Strength”로 이용한다. 또한 각 대영역에 대한 “성별-교육정도별” 8개 범주의 실업자 총계는 해당 대영역을 포함하고 있는 부차그룹의

“성별-연령대별” 8개 범주의 실업자 총계 추정값들을 “Borrow Strength” 로 적용하여 추정한다.

6.2 특·광역시 및 도지역 집락화 방안

경제활동인구조사에서 대영역인 특·광역시 및 도지역들에 대한 표본 조사구 수의 분포를 살펴보면 다음 <표6.1>과 같다.

<표6.1> 특·광역시 및 도지역 표본조사구 수 분포

시도	주민등록인구	표본조사구	
		조사구 수	구성비(%)
서울	10,331,244	207	12.71
부산	3,786,033	117	7.18
대구	2,539,587	90	5.52
인천	2,581,557	108	6.63
광주	1,387,360	81	4.97
대전	1,408,809	81	4.97
울산	1,060,378	63	3.87
경기도	9,612,036	198	12.15
강원도	1,556,904	90	5.52
충청북도	1,504,518	81	4.97
충청남도	1,928,088	90	5.52
전라북도	2,013,923	81	4.97
전라남도	2,104,052	90	5.52
경상북도	2,802,597	108	6.63
경상남도	3,124,123	99	6.08
제주도	547,964	45	2.76
합계	48,289,173	1,629	100.00

* 주민등록인구는 2001년 12월 기준

16개 대영역의 전체 표본조사구 수는 총 1,629개이다. 표본조사구 수가 가장 적게 배정된 지역은 제주도로써 45개이고 전체 표본조사구 중 약 2.76%에 해당한다. 표본조사구 수가 가장 많은 지역은 서울특별시로써

207개이고 전체 표본조사구 중 약 12.71%에 해당한다. 여기에서 대영역별 표본조사구 수의 배정은 대영역에 대한 실업통계 추정값들의 허용오차를 고려하여 결정된다. <표6.1>은 대영역별 실업자 총계와 같은 단순항목 통계를 생산하기 위해 설계된 표본조사구 수 배정 결과이다.

경제활동인구조사 자료를 이용하여 대영역 내의 세부영역인 “성별-연령대별” 또는 “성별-교육정도별” 범주들에 대한 실업자 총계 추정을 고려해 보자. 이때 “성별-연령대별” 또는 “성별-교육정도별” 세부범주들은 표본설계에 반영된 관심영역이 아니기 때문에 세부영역별 실업자 총계 추정값들은 매우 큰 추정오차를 동반할 수밖에 없다. 이러한 문제점을 해소하기 위한 하나의 방법은 유사 인근지역의 정보를 활용하는 방법이다. 대영역들을 실업률이 유사한 동질적인 그룹들로 묶어 간접적으로 충분한 표본수를 확보한 후 각 그룹에 대한 “성별-연령대별” 또는 “성별-교육정도별” 세부범주에 대한 실업자 총계를 추정한 후 이러한 정보를 그룹 내의 대영역 세부범주 항목통계 추정에 이용하는 방법이다.

대영역들을 실업률이 유사한 동질적인 그룹들로 집락화하는 방법을 소개하면 다음과 같다. 집락분석에 이용된 자료는 보다 정확한 분석을 위해 2000년 「인구주택 총 조사」 자료를 이용하였다. 「인구주택 총 조사」 자료에서 대영역별로 “성별(남, 여)-연령대별(15-29세, 30-39세, 40-49세, 50세 이상)-교육정도별(초등졸 이하, 중졸, 고졸, 대졸 이상)”의 총 32개 범주의 실업률 X_1, X_2, \dots, X_{32} 를 계산하여 집락화를 위한 연구변수들로 이용하였다. 이러한 자료로 구성된 데이터 셀을 “temp.txt”라 하자.

<u>temp.txt</u>	AREA	X_1	X_2	...	X_{32}
	1	x_{11}	x_{12}	...	$x_{1,32}$
	2	x_{21}	x_{22}	...	$x_{2,32}$
	⋮				
	16	$x_{16,1}$	$x_{16,2}$...	$x_{16,32}$

여기에서 x_{ij} 는 i 번째 대영역에서 j 범주의 실업률을 나타낸다.

SAS의 FASTCLUS 프로시저를 이용하여 집락분석 프로그램을 작성하면 다음과 같다.

집락화 프로그램

```
DATA temp;
  INFILE 'A:\temp.txt';
  INPUT AREA X1-X32;
PROC FASTCLUS DATA=temp OUT=result
  MAXCLUSTERS=4 MAXITER=10 CONV=0.02;
  VAR X1-X32;
PROC PRINT DATA=result;
RUN;
```

FASTCLUS 처리절차는 유클리드 거리를 기초로 반복적 알고리즘을 통해 상호 배반적 집락을 형성시키며 대량의 자료를 연구의 목적에 적합한 동질적인 그룹들의 개수(MAXCLUSTERS=n)로 집락화 하는데 상당히 효과적인 방법으로 알려져 있다.

FASTCLUS 프로시저를 이용한 집락분석결과는 다음 <표6.2>와 같다.

<표6.2> 집락분석 결과

() : 표본조사구 수

집락1	집락2	집락3	집락4
경기도(198), 울산시(63)	서울시(207), 부산시(117), 인천시(108), 대전시(81), 대구시(90), 광주시(81)	충청남도(90) 경상남도(99), 경상북도(108), 제주도(45)	강원도(90), 충청북도(81), 전라남도(90), 전라북도(81)

전체 시지역들 중 “울산시”를 제외한 나머지 지역들인 “서울, 부산, 인천, 대전, 대구, 광주시” 지역들은 “성별-교육정도별-연령대별” 실업률이 유사한 동일한 집락으로 구분되었다. 도지역들에 대해서는 “충청남도, 경상남도, 경상북도, 제주도”가 하나의 집락으로, “강원도, 충청북도, 전라남도, 전라북도”가 또 다른 하나의 집락으로 구분되었다. 특이한 사항은 “경기도”와 “울산시”가 실업률이 유사한 동일한 집락으로 구분되었다는 점이다. “경기도”와 “울산시”를 제외한 나머지 지역들의 집락화 경향을 살펴보면 시지역들과 군지역들이 상호 배반적인 집락들로 구분되는 현상을 나타낸다.

이상의 집락분석 결과를 토대로 “울산시”를 유사집락2에 포함시키고, “경기도”를 유사집락3에 포함시켜 16개 대영역을 한개의 시지역들에 대한 집락과 2개의 군지역들에 대한 집락으로 구분한 후 이러한 집락들에 대한 “성별-연령대별-교육정도별” 세부범주의 실업자 총계를 경제활동인구조사 자료에서 추정된 후 집락 내의 각 대영역별 세부범주들에 대한 실업자 총계 추정 시 “Borrow Strength”로 활용하고자 한다.

6.3 소지역 추정법 적용

대영역별 실업자 총계 추정범주는 “성별-연령대별” 8개 범주와 “성별-교육정도별” 8개 범주로써 세부범주를 명시하면 다음과 같다.

성별-연령대별 추정범주(8개 범주)

- 남자 15-29세 실업자 총계
- 남자 30-39세 실업자 총계
- 남자 40-49세 실업자 총계
- 남자 50세 이상 실업자 총계
- 여자 15-29세 실업자 총계
- 여자 30-39세 실업자 총계
- 여자 40-49세 실업자 총계

- 여자 50세 이상 실업자 총계

성별-교육정도별 추정범주(8개 범주)

- 남자 초등학교 졸 이하 실업자 총계
- 남자 중졸 실업자 총계
- 남자 고졸 실업자 총계
- 남자 대졸 이상 실업자 총계
- 여자 초등학교 졸 이하 실업자 총계
- 여자 중졸 실업자 총계
- 여자 고졸 실업자 총계
- 여자 대졸 이상 실업자 총계

제 5장에서 소개한 소지역 추정방법을 적용하여 대영역 내에서 “성별-연령대별” 8개 범주에 대한 실업자 총계 및 “성별-교육정도별” 8개 범주의 실업자 총계를 추정하고자 한다. 편의상 “성별-연령대별” 8개 범주에 대한 실업자 총계 추정방법을 구체적으로 소개하기로 한다.

대영역의 “성별-연령대별” 세부범주 i 의 실업자 총계를 추정하기 위한 직접추정공식은 다음과 같다.

$$\hat{Y}_{i.} = \sum_{h=1}^{n_i} M_i Y_{ih}, \quad i=1,2,\dots,8; h=1,2,\dots,n_i \quad (6.1)$$

여기에서 n_i 는 경제활동인구조사에서 세부범주 i 의 표본조사구 수, Y_{ih} 는 세부범주 i 의 표본 조사구에서 조사한 실업자 수를 나타낸다. 승수 $M_i = \hat{X}_{i.} / X_{i.}$ 은 $\hat{Y}_{i.}$ 이 불편추정량이 되도록 산정 한다. 여기에서 $\hat{X}_{i.}$ 은 대영역 내에서 세부범주 i 에 대한 15세 이상의 상주 추계인구를 나타내며, $X_{i.}$ 는 경제활동인구조사에서 조사된 15세 이상의 조사인구를 나타낸다.

이때 직접추정량 $\hat{Y}_{i.}$ 의 분산은

$$Var(\hat{Y}_{i.}) = Var\left(\sum_{h=1}^{n_i} M_i Y_{ih}\right) = M_i^2 Var\left(\sum_{h=1}^{n_i} Y_{ih}\right)$$

로 주어진다. 통계청의 연속차 분산추정공식을 적용하면 직접추정량의 분산은

$$\widehat{Var}(\hat{Y}_{i.}) = M_i^2 \zeta_i \sum_{h=1}^{n_i} U_{ih}^2 \quad (6.2)$$

으로 추정될 수 있다.

여기에서

$$U_{ih} = d Y_{ih} - \rho_i \cdot d X_{ih} ,$$

$$d Y_{ih} = Y_{ih} - Y_{i,h+1} ,$$

$$d X_{ih} = X_{ih} - X_{i,h+1} ,$$

$$\rho_i = Y_{i.} / X_{i.} ,$$

$$\zeta_i = [1 - n_i / (10N_i)] n_i / [2(n_i - 1)] ,$$

N_i = 표본추출틀의 조사구 수.

대영역의 “성별(남, 여)-연령대별(15-29세, 30-39세, 40-49세, 50세 이상)” 세부범주 i 에 대한 합성추정량 $\hat{Y}_{i.}^c$ 는 세부범주 i 와 특성이 유사한 인근 대영역의 정보를 추정에 이용하는 간접적인 설계 기반 추정량이다.

우선 16개 대영역들을 실업률의 특성이 유사한 동질적인 3개의 집락으로 분할하고, 경제활동인구조사 자료에서 각 집락별로 “성별(남, 여)-교육정도별(초등졸 이하, 중졸, 고졸, 대졸 이상)” 세부범주에 대한 실업자 총계를 추정한다. 대영역의 “성별-연령대별” 세부범주에 대한 실업자 총계의 합성추정값은 대영역을 포함하고 있는 해당 집락의 “성별-교육정도별” 세부범주의 실업자 총계 추정값들을 이용하여 추정한다. 한편, 대영역의 “성별-교육정도별” 세부범주에 대한 실업자 총계의 합성추정값은 대영역을 포함하고 있는 해당 집락의 “성별-연령대별” 세부범주의 실업자 총계

추정값들을 이용하여 추정한다. 앞서 집락화 방안에서 언급하였듯이 3개의 집락은 각각 “서울, 부산, 인천, 대전, 대구, 울산, 광주” 집락, “경기, 충남, 경남, 경북, 제주” 집락과 “강원, 충북, 전남, 전북” 집락이다.

16개의 대영역들이 3개의 집락으로 분할되었다고 하자. 또한 각각의 집락은 I_k ($k=1, 2, 3$)개의 대영역들을 포함한다고 하자. 이때 각 집락별로 집락 내에 있는 대영역들의 “성별-연령대별” 세부범주에 대한 실업자 총계를 추정하고자 한다. 대영역의 i 번째 “성별-연령대별” 세부범주에 대한 실업자 총계의 합성추정량 \hat{Y}_i^s 는 다음과 같이 주어질 수 있다.

$$\hat{Y}_i^s = \sum_{j=1}^4 \frac{j\hat{P}_i}{j\hat{X}_i} j\hat{Y}_{i,dir}, \quad i=1, 2, \dots, I_k, \quad (6.3)$$

여기에서 j 는 “성별(남, 여)-교육정도별(초등졸 이하, 중졸, 고졸, 대졸 이상)” 세부범주를 나타낸다. $j\hat{P}_i$ 는 행정자료로부터 산정된 i 번째 “성별-연령대별” 범주에서 j 번째 “성별-교육정도별” 범주의 상주추정인구를 나타낸다. $j\hat{X}_i$ 과 $j\hat{Y}_{i,dir}$ 은

$$j\hat{X}_i = \sum_{u=1}^4 j\hat{X}_u,$$

$$j\hat{Y}_{i,dir} = \sum_{l=1}^4 \sum_{h=1}^{n_i} jM_{ul} Y_{ih}$$

로 주어진다. $j\hat{X}_i$ 는 i 번째 “성별-연령대별” 범주와 j 번째 “성별-교육정도별” 범주에 대해서 집락 내의 I_k 개의 대영역들의 상주추정인구 합을 나타내며, $j\hat{Y}_{i,dir}$ 는 실업자 총계에 대한 직접추정값을 나타낸다. 승수 jM_{ul} 은 상주추정인구 $j\hat{X}_u$ 와 경찰자료로부터 조사된 $jX_u = \sum_{h=1}^{n_u} jX_{uh}$ 에 의해 결정되며 $jM_{ul} = j\hat{X}_u / jX_u$ 로 주어진다.

합성추정량 \hat{Y}_i^s 에 대한 정확도의 측도로써 잭나이프 평균제공오차 추

정량을 고려할 수 있으며 제 5장에 언급되었던 추정절차를 적용하여 산정할 수 있다. i 번째 “성별-연령대별” 범주에서 표본조사구를 하나씩 제외하고 반복적으로 추정된 합성추정값들을 $\hat{Y}_{i.}^s(l)$ ($l=1,2,\dots,n_i$)라 했을 때 잭나이프 추정공식들은 다음과 같이 주어진다.

$$mse_J(\hat{Y}_{i.}^s) = \widehat{Var}_J(\hat{Y}_{i.}^s) + [\widehat{Bias}_J(\hat{Y}_{i.}^s)]^2, \quad i=1,2,\dots,8$$

여기에서 편향과 분산은

$$\widehat{Bias}_J(\hat{Y}_{i.}^s) = (n_i - 1)[\hat{Y}_{i.}^s(\cdot) - \hat{Y}_{i.}^s],$$

$$\widehat{Var}_J(\hat{Y}_{i.}^s) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} [\hat{Y}_{i.}^s(h) - \hat{Y}_{i.}^s(\cdot)]^2$$

로 추정되며, $\hat{Y}_{i.}^s(\cdot)$ 는

$$\hat{Y}_{i.}^s(\cdot) = (\hat{Y}_{i.}^s(1) + \hat{Y}_{i.}^s(2) + \dots + \hat{Y}_{i.}^s(n_i)) / n_i$$

로 주어진다.

대영역의 i 번째 “성별-연령대별” 범주의 실업자 총계에 대한 복합추정량 $\hat{Y}_{i.}^C$ 는 직접추정량 $\hat{Y}_{i.}$ 와 합성추정량 $\hat{Y}_{i.}^s$ 의 가중평균으로 표현되며 다음과 같이 주어진다.

$$\hat{Y}_{i.}^C = w_i \hat{Y}_{i.} + (1 - w_i) \hat{Y}_{i.}^s, \quad i=1,2,\dots,8 \quad (6.4)$$

여기에서 가중치 w_i 는 0과 1사이의 값을 취하며 경제활동인구조사 자료로부터 추정되어야 할 값이다. (6.4)식에서 $Cov(\hat{Y}_{i.}, \hat{Y}_{i.}^s) = 0$ 를 가정한다면, 복합추정량 $\hat{Y}_{i.}^C$ 의 평균제곱오차를 최소화하는 가중치는

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

에 의해 결정된다. 따라서 (6.4)식의 복합추정량은 경제활동인구조사 자료로부터 추정된 최적가중치 $\hat{w}_{i(opt)}$ 를 이용하여 다음 식으로부터 추정될 수 있다.

$$\hat{Y}_{i.}^c = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s \quad (6.5)$$

복합추정값들의 잭나이프 평균제곱오차는 다음 추정식으로부터 산정할 수 있다.

$$mse_J(\hat{Y}_{i.}^c) = \widehat{Var}_J(\hat{Y}_{i.}^c) + [\widehat{Bias}_J(\hat{Y}_{i.}^c)]^2$$

여기에서

$$\widehat{Var}_J(\hat{Y}_{i.}^c) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} \left[\hat{Y}_{i.}^c(h) - \frac{1}{n_i} \sum_{l=1}^{n_i} \hat{Y}_{i.}^c(l) \right]^2,$$

$$\widehat{Bias}_J(\hat{Y}_{i.}^c) = (n_i - 1) \left[\frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}^c(h) - \hat{Y}_{i.}^c \right],$$

$$\hat{Y}_{i.}^c(h) = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^s(h)$$

로 주어진다.

6.4 SAS 프로그램 알고리즘

집락 내 대영역의 “성별(남, 여)-연령대별(15-29세, 30-39세, 40-49세, 50세 이상)” 8개 범주에 대한 실업자 총계 추정절차를 소개한다. 편의상 집락 내 대영역의 수를 2개 지역으로 가정하여 설명하기로 한다.

(Step1) 직접추정값 계산

(1) “성별-연령대별” Data Set 작성 및 범주별 직접추정값 계산

대영역	표본 조사구	남자15-29세 실업자	남자30-39세 실업자	남자40-49세 실업자	남자50세이상 실업자	여자15-29세 실업자	여자30-39세 실업자	여자40-49세 실업자	여자50세이상 실업자
A	1 2 ⋮ n_A	Y_{11} Y_{12} ⋮ Y_{1n_A}	Y_{21} Y_{22} ⋮ Y_{2n_A}	⋮	⋮	⋮	⋮	⋮	Y_{81} Y_{82} ⋮ Y_{8n_A}
B	1 2 ⋮ n_B	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

* 대영역 A의 i 번째 “성별-연령대별” 범주의 표본조사구 수는 대영역 A의 표본조사구와 동일함($n_i = n_A$).

* 식(6.1)과 식(6.2)로부터 대영역 A와 B의 i 번째 “성별-연령대별” 범주의 직접추정값과 추정값의 분산을 계산함.

(2) “성별-교육정도별” Data Set 작성 및 범주별 직접추정값 계산

대영역	표본 조사구	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	1 2 ⋮ n_A	Y_{11} Y_{12} ⋮ Y_{1n_A}	Y_{21} Y_{22} ⋮ Y_{2n_A}	⋮	⋮	⋮	⋮	⋮	Y_{81} Y_{82} ⋮ Y_{8n_A}
B	1 2 ⋮ n_B	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

* 식(6.1)과 식(6.2)로부터 대영역 A의 i 번째 “성별-교육정도별” 범주의 직접추정값, 추정값의 분산, CV값들을 계산함.

(Step2) 합성추정값 계산

(1) “성별-연령대별-교육정도별” 상주추정인구, 상주조사인구와 실업자 수의 Data Set을 작성한다. 상주추정인구는 통계청의 추계인구를 이용하고, 상주조사인구와 실업자 수는 경찰자료로부터 계산한다.

<성별-연령대별-교육정도별 상주추정인구(통계청)>

대영역	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남자15-29세	Am11	Am12	Am13	Am14				
	남자30-39세	Am21	Am22	Am23	Am24				
	남자40-49세	Am31	Am32	Am33	Am34				
	남자50세+	Am41	Am42	Am43	Am44				
	여자15-29세					Af15	Af16	Af17	Af18
	여자30-39세					Af25	Af26	Af27	Af28
	여자40-49세					Af35	Af36	Af37	Af38
	여자50+					Af45	Af46	Af47	Af48
B	남자15-29세	Bm11	Bm12	Bm13	Bm14				
	남자30-39세	Bm21	Bm22	Bm23	Bm24				
	남자40-49세	Bm31	Bm32	Bm33	Bm34				
	남자50세+	Bm41	Bm42	Bm43	Bm44				
	여자15-29세					Bf15	Bf16	Bf17	Bf18
	여자30-39세					Bf25	Bf26	Bf27	Bf28
	여자40-49세					Bf35	Bf36	Bf37	Bf38
	여자50+					Bf45	Bf46	Bf47	Bf48

<성별-연령대별-교육정도별 상주조사인구(경찰자료)>

대영역	세부범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남자15-29세	aM11	aM12	aM13	aM14				
	남자30-39세	aM21	aM22	aM23	aM24				
	남자40-49세	aM31	aM32	aM33	aM34				
	남자50세+	aM41	aM42	aM43	aM44				
	여자15-29세					aF15	aF16	aF17	aF18
B	남자15-29세	bM11	bM12	bM13	bM14				
	남자30-39세	bM21	bM22	bM23	bM24				
	남자40-49세	bM31	bM32	bM33	bM34				
	남자50세+	bM41	bM42	bM43	bM44				
	여자15-29세					bF15	bF16	bF17	bF18
여자30-39세					bF25	bF26	bF27	bF28	
여자40-49세					bF35	bF36	bF37	bF38	
여자50+					bF45	bF46	bF47	bF48	

<성별-연령대별-교육정도별 실업자 수(경찰자료)>

대영역	세부범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남자15-29세	aMu11	aMu12	aMu13	aMu14				
	남자30-39세	aMu21	aMu22	aMu23	aMu24				
	남자40-49세	aMu31	aMu32	aMu33	aMu34				
	남자50세+	aMu41	aMu42	aMu43	aMu44				
	여자15-29세					aFu15	aFu16	aFu17	aFu18
B	남자15-29세	bMu11	bMu12	bMu13	bMu14				
	남자30-39세	bMu21	bMu22	bMu23	bMu24				
	남자40-49세	bMu31	bMu32	bMu33	bMu34				
	남자50세+	bMu41	bMu42	bMu43	bMu44				
	여자15-29세					bFu15	bFu16	bFu17	bFu18
여자30-39세					bFu25	bFu26	bFu27	bFu28	
여자40-49세					bFu35	bFu36	bFu37	bFu38	
여자50+					bFu45	bFu46	bFu47	bFu48	

(2) 경찰자료로부터 “성별-연령대별-교육정도별” 실업자 총계를 추정한다. 범주별 추정결과는 다음과 같다.

대영역	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남자15-29세	AM11	AM12	AM13	AM14				
	남자30-39세	AM21	AM22	AM23	AM24				
	남자40-49세	AM31	AM32	AM33	AM34				
	남자50세+	AM41	AM42	AM43	AM44				
	여자15-29세					AF15	AF16	AF17	AF18
	여자30-39세					AF25	AF26	AF27	AF28
	여자40-49세					AF35	AF36	AF37	AF38
	여자50+					AF45	AF46	AF47	AF48
B	남자15-29세	BM11	BM12	BM13	BM14				
	남자30-39세	BM21	BM22	BM23	BM24				
	남자40-49세	BM31	BM32	BM33	BM34				
	남자50세+	BM41	BM42	BM43	BM44				
	여자15-29세					BF15	BF16	BF17	BF18
	여자30-39세					BF25	BF26	BF27	BF28
	여자40-49세					BF35	BF36	BF37	BF38
	여자50+					BF45	BF46	BF47	BF48

* 대영역 A의 “남자 15-29세”-“남자 초등졸” 범주의 실업자 총계 (=AM11)의 계산식은 다음과 같다.

$$AM11 = \frac{Am11}{aM11} \times aMu11$$

나머지 범주의 실업자 총계도 같은 방법으로 계산한다.

(3) 집락에 대한 “성별-연령대별-교육정도별” 실업자 총계 계산

집락	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
C= A+B	남자15-29세	CM11	CM12	CM13	CM14				
	남자30-39세	CM21	CM22	CM23	CM24				
	남자40-49세	CM31	CM32	CM33	CM34				
	남자50세+	CM41	CM42	CM43	CM44				
	여자15-29세					CF15	CF16	CF17	CF18
	여자30-39세					CF25	CF26	CF27	CF28
	여자40-49세					CF35	CF36	CF37	CF38
	여자50+					CF45	CF46	CF47	CF48

* 집락C(=A+B)의 “남자15-29세”-“남자 초등졸” 범주의 실업자 총계 (=CM11)는 $CM11 = AM11 + BM11$ 으로 계산한다. 나머지 범주도 같은 방법으로 계산한다.

(4) 대영역의 “성별-연령대별” 범주의 합성추정값 계산

대영역	세부범주	합성추정값	대영역	세부범주	합성추정값
A	남자15-29세	\hat{Y}_{A1}^S	B	남자15-29세	\hat{Y}_{B1}^S
	남자30-39세	\hat{Y}_{A2}^S		남자30-39세	\hat{Y}_{B2}^S
	남자40-49세	\hat{Y}_{A3}^S		남자40-49세	\hat{Y}_{B3}^S
	남자50세+	\hat{Y}_{A4}^S		남자50세+	\hat{Y}_{B4}^S
	여자15-29세	\hat{Y}_{A5}^S		여자15-29세	\hat{Y}_{B5}^S
	여자30-39세	\hat{Y}_{A6}^S		여자30-39세	\hat{Y}_{B6}^S
	여자40-49세	\hat{Y}_{A7}^S		여자40-49세	\hat{Y}_{B7}^S
	여자50+	\hat{Y}_{A8}^S		여자50+	\hat{Y}_{B8}^S

* 대영역A의 “남자 15-29세” 범주의 실업자 총계에 대한 합성추정값 \hat{Y}_{A1}^S 는 다음과 같이 계산한다.

$$\hat{Y}_{A1}^s = \left(\frac{Am11}{Am11+Bm11} \times CM11\right) + \left(\frac{Am12}{Am12+Bm12} \times CM12\right) \\ + \left(\frac{Am13}{Am13+Bm13} \times CM13\right) + \left(\frac{Am14}{Am14+Bm14} \times CM14\right)$$

* 대영역A의 “여자 15-29세” 범주의 실업자 총계에 대한 합성추정값 \hat{Y}_{A5}^s 는 다음과 같이 계산한다.

$$\hat{Y}_{A5}^s = \left(\frac{Afl5}{Afl5+Bfl5} \times CF15\right) + \left(\frac{Afl6}{Afl6+Bfl6} \times CF16\right) \\ + \left(\frac{Afl7}{Afl7+Bfl7} \times CF17\right) + \left(\frac{Afl8}{Afl8+Bfl8} \times CF18\right)$$

* 나머지 범주에 대해서도 같은 방법으로 계산한다.

(Step3) 합성추정값의 잭나이프 Bias, 분산, 평균제곱오차 계산

(1) 대영역A의 “성별-연령대별” 범주에서 반복 합성추정량들을 생성한다. 다음의 과정을 반복하면 대영역A의 i 번째 “성별-연령대별” 범주의 표본조사구 수(= $n_i = n_A$) 만큼 반복합성추정량들이 생성된다. 대영역A에서 l 번째 표본조사구를 제외하고 나머지 표본조사구들을 이용하여 i 번째 “성별-연령대별” 범주의 새로운 합성추정값 (= $\hat{Y}_{Ai}^s(l)$, $l=1, 2, \dots, n_i$, $i=1, \dots, 8$)을 다음 절차에 의해 반복적으로 생성한다.

① 대영역A의 l 번째 표본조사구를 제외하고 다음 표와 같이 범주별 상주조사인구를 다시 카운트한다. 여기에서 $aM11(l)$ 은 l 번째 표본조사구를 제외하고 카운트된 “남자 15-29세”-“남자 초등졸” 범주의 상주조사인구를 나타낸다.

<성별-연령대별-교육정도별 상주조사인구(경찰자료)>

대영역	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남자15-29세	aM11(l)	aM12(l)	aM13(l)	aM14(l)				
	남자30-39세	aM21(l)	aM22(l)	aM23(l)	aM24(l)				
	남자40-49세	aM31(l)	aM32(l)	aM33(l)	aM34(l)				
	남자50세+	aM41(l)	aM42(l)	aM43(l)	aM44(l)				
	여자15-29세					aF15(l)	aF16(l)	aF17(l)	aF18(l)
	여자30-39세					aF25(l)	aF26(l)	aF27(l)	aF28(l)
	여자40-49세					aF35(l)	aF36(l)	aF37(l)	aF38(l)
여자50+					aF45(l)	aF46(l)	aF47(l)	aF48(l)	
B	남자15-29세	bM11	bM12	bM13	bM14				
	남자30-39세	bM21	bM22	bM23	bM24				
	남자40-49세	bM31	bM32	bM33	bM34				
	남자50세+	bM41	bM42	bM43	bM44				
	여자15-29세					bF15	bF16	bF17	bF18
	여자30-39세					bF25	bF26	bF27	bF28
	여자40-49세					bF35	bF36	bF37	bF38
여자50+					bF45	bF46	bF47	bF48	

② 대영역A의 l번째 표본조사구를 제외하고 범주별 실업자 수를 다시 카운트한다. 다음 표에서 aMu11(l)은 대영역A에서 l번째 표본조사구를 제외하고 카운트된 “남자15-29세”-“남자 초등졸” 실업자 수를 나타낸다.

<성별-연령대별-교육정도별 실업자 수(경찰자료)>

대영역	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남15-29세	aMu11(l)	aMu12(l)	aMu13(l)	aMu14(l)				
	남30-39세	aMu21(l)	aMu22(l)	aMu23(l)	aMu24(l)				
	남40-49세	aMu31(l)	aMu32(l)	aMu33(l)	aMu34(l)				
	남50세+	aMu41(l)	aMu42(l)	aMu43(l)	aMu44(l)				
	여15-29세					aFu15(l)	aFu16(l)	aFu17(l)	aFu18(l)
	여30-39세					aFu25(l)	aFu26(l)	aFu27(l)	aFu28(l)
	여40-49세					aFu35(l)	aFu36(l)	aFu37(l)	aFu38(l)
여50+					aFu45(l)	aFu46(l)	aFu47(l)	aFu48(l)	
B	남15-29세	bMu11	bMu12	bMu13	bMu14				
	남30-39세	bMu21	bMu22	bMu23	bMu24				
	남40-49세	bMu31	bMu32	bMu33	bMu34				
	남50세+	bMu41	bMu42	bMu43	bMu44				
	여15-29세					bFu15	bFu16	bFu17	bFu18
	여30-39세					bFu25	bFu26	bFu27	bFu28
	여40-49세					bFu35	bFu36	bFu37	bFu38
여50+					bFu45	bFu46	bFu47	bFu48	

③ 대영역A의 범주별 실업자 총계를 추정한다. 범주별 추정결과는 다음과 같다.

영역	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
A	남15-29세	AM11(l)	AM12(l)	AM13(l)	AM14(l)				
	남30-39세	AM21(l)	AM22(l)	AM23(l)	AM24(l)				
	남40-49세	AM31(l)	AM32(l)	AM33(l)	AM34(l)				
	남50세+	AM41(l)	AM42(l)	AM43(l)	AM44(l)				
	여15-29세					AF15(l)	AF16(l)	AF17(l)	AF18(l)
	여30-39세					AF25(l)	AF26(l)	AF27(l)	AF28(l)
	여40-49세					AF35(l)	AF36(l)	AF37(l)	AF38(l)
	여50+					AF45(l)	AF46(l)	AF47(l)	AF48(l)
B	남15-29세	BM11	BM12	BM13	BM14				
	남30-39세	BM21	BM22	BM23	BM24				
	남40-49세	BM31	BM32	BM33	BM34				
	남50세+	BM41	BM42	BM43	BM44				
	여15-29세					BF15	BF16	BF17	BF18
	여30-39세					BF25	BF26	BF27	BF28
	여40-49세					BF35	BF36	BF37	BF38
	여50+					BF45	BF46	BF47	BF48

* 대영역 A의 “남자 15-29세”-“남자 초등졸” 범주의 실업자 총계 (=AM11(l))의 계산식은 다음과 같다.

$$AM11(l) = \frac{Am11}{aM11(l)} \times aMu11(l), \quad l=1,2,\dots,n_A$$

대영역A의 나머지 범주의 실업자 총계도 같은 방법으로 계산한다.

④ 집락에 대한 범주별 실업자 총계를 계산한다.

집락	세부 범주	남자 초등졸	남자 중졸	남자 고졸	남자 대졸	여자 초등졸	여자 중졸	여자 고졸	여자 대졸
C= A+B	남15-29세	CM11(l)	CM12(l)	CM13(l)	CM14(l)				
	남30-39세	CM21(l)	CM22(l)	CM23(l)	CM24(l)				
	남40-49세	CM31(l)	CM32(l)	CM33(l)	CM34(l)				
	남50세+	CM41(l)	CM42(l)	CM43(l)	CM44(l)				
	여15-29세					CF15(l)	CF16(l)	CF17(l)	CF18(l)
	여30-39세					CF25(l)	CF26(l)	CF27(l)	CF28(l)
	여40-49세					CF35(l)	CF36(l)	CF37(l)	CF38(l)
	여50+					CF45(l)	CF46(l)	CF47(l)	CF48(l)

* 집락C(=A+B)의 “남자15-29세”-“남자 초등졸” 범주의 실업자 총계 (=CM11(l))는

$$CM11(l) = AM11(l) + BM11, l=1,2,\dots,n_A$$

으로 계산한다. 나머지 범주도 같은 방법으로 계산한다.

⑤ 대영역A의 “성별-연령대별” 범주의 반복 합성추정값 계산

대영역	세부범주	합성추정값	대영역	세부범주	합성추정값
A	남자15-29세	$\hat{Y}_{A1}^s(l)$	B	남자15-29세	$\hat{Y}_{B1}^s(l)$
	남자30-39세	$\hat{Y}_{A2}^s(l)$		남자30-39세	$\hat{Y}_{B2}^s(l)$
	남자40-49세	$\hat{Y}_{A3}^s(l)$		남자40-49세	$\hat{Y}_{B3}^s(l)$
	남자50세+	$\hat{Y}_{A4}^s(l)$		남자50세+	$\hat{Y}_{B4}^s(l)$
	여자15-29세	$\hat{Y}_{A5}^s(l)$		여자15-29세	$\hat{Y}_{B5}^s(l)$
	여자30-39세	$\hat{Y}_{A6}^s(l)$		여자30-39세	$\hat{Y}_{B6}^s(l)$
	여자40-49세	$\hat{Y}_{A7}^s(l)$		여자40-49세	$\hat{Y}_{B7}^s(l)$
	여자50+	$\hat{Y}_{A8}^s(l)$		여자50+	$\hat{Y}_{B8}^s(l)$

* 대영역A의 “남자 15-29세” 범주의 실업자 총계에 대한 반복 합성추정값 $\hat{Y}_{A1}^s(l)$, $l=1,2,\dots,n_A$ 는 다음과 같이 계산한다.

$$\begin{aligned} \hat{Y}_{A1}^s(l) = & \left(\frac{Am11}{Am11 + Bm11} \times CM11(l) \right) + \left(\frac{Am12}{Am12 + Bm12} \times CM12(l) \right) \\ & + \left(\frac{Am13}{Am13 + Bm13} \times CM13(l) \right) + \left(\frac{Am14}{Am14 + Bm14} \times CM14(l) \right) \end{aligned}$$

* 대영역A의 “여자 15-29세” 범주의 실업자 총계에 대한 합성추정값 $\hat{Y}_{A5}^s(l)$, $l=1,2,\dots,n_A$ 는 다음과 같이 계산한다.

$$\begin{aligned} \hat{Y}_{A5}^s = & \left(\frac{Af15}{Af15 + Bf15} \times CF15(l) \right) + \left(\frac{Af16}{Af16 + Bf16} \times CF16(l) \right) \\ & + \left(\frac{Af17}{Af17 + Bf17} \times CF17(l) \right) + \left(\frac{Af18}{Af18 + Bf18} \times CF18(l) \right) \end{aligned}$$

* 대영역B에 대한 반복합성추정값도 같은 방법으로 위의 절차 ①~④로부터 계산할 수 있다.

⑥ 대영역A의 범주별 반복 합성추정값들을 이용하여 잭나이프 Bias, 분산, 평균제곱오차와 CV값을 계산한다. 추정식은 다음과 같다.

$$\widehat{Bias}_J(\hat{Y}_{Ai}^S) = (n_i - 1)[\hat{Y}_{Ai}^S(\cdot) - \hat{Y}_{Ai}^S], \quad i = 1, 2, \dots, 8$$

$$\widehat{Var}_J(\hat{Y}_{Ai}^S) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} [\hat{Y}_{Ai}^S(h) - \hat{Y}_{Ai}^S(\cdot)]^2,$$

$$mse_J(\hat{Y}_{Ai}^S) = \widehat{Var}_J(\hat{Y}_{Ai}^S) + [\widehat{Bias}_J(\hat{Y}_{Ai}^S)]^2,$$

$$CV(\hat{Y}_{Ai}^S) = \frac{\sqrt{mse_J(\hat{Y}_{Ai}^S)}}{\hat{Y}_{Ai}^S} \times 100,$$

여기에서 $\hat{Y}_{Ai}^S(\cdot) = (\hat{Y}_{Ai}^S(1) + \hat{Y}_{Ai}^S(2) + \dots + \hat{Y}_{Ai}^S(n_i)) / n_i$ 이다.

(Step4) 복합추정값 계산

(1) 대영역의 “성별-연령대별” 범주들의 최적가중치를 계산한다.

대영역	남자15 -29세	남자30 -39세	남자40 -49세	남자50 세이상	여자15 -29세	여자30 -39세	여자40 -49세	여자50 세이상
A	\hat{w}_{A1}	\hat{w}_{A2}	\hat{w}_{A3}	\hat{w}_{A4}	\hat{w}_{A5}	\hat{w}_{A6}	\hat{w}_{A7}	\hat{w}_{A8}
B	\hat{w}_{B1}	\hat{w}_{B2}	\hat{w}_{B3}	\hat{w}_{B4}	\hat{w}_{B5}	\hat{w}_{B6}	\hat{w}_{B7}	\hat{w}_{B8}

* 최적가중치 산출공식:

$$\hat{w}_{Ai} = \frac{mse_J(\hat{Y}_{Ai}^S)}{mse_J(\hat{Y}_{Ai}^S) + \widehat{Var}(\hat{Y}_{Ai}^S)}, \quad i = 1, 2, \dots, 8,$$

$$\hat{w}_{Bi} = \frac{mse_J(\hat{Y}_{Bi}^S)}{mse_J(\hat{Y}_{Bi}^S) + \widehat{Var}(\hat{Y}_{Bi}^S)}$$

여기에서 $\widehat{Var}(\hat{Y}_{Ai}^S)$ 와 $\widehat{Var}(\hat{Y}_{Bi}^S)$ 는 (Step1)의 직접추정값의 분산임.

(2) 대영역의 “성별-연령대별” 범주의 복합추정값을 계산한다.

대영역	남자15-29세	남자30-39세	남자40-49세	남자50세이상	여자15-29세	여자30-39세	여자40-49세	여자50세이상
A	\hat{Y}_{A1}^C	\hat{Y}_{A2}^C	\hat{Y}_{A3}^C	\hat{Y}_{A4}^C	\hat{Y}_{A5}^C	\hat{Y}_{A6}^C	\hat{Y}_{A7}^C	\hat{Y}_{A8}^C
B	\hat{Y}_{B1}^C	\hat{Y}_{B2}^C	\hat{Y}_{B3}^C	\hat{Y}_{B4}^C	\hat{Y}_{B5}^C	\hat{Y}_{B6}^C	\hat{Y}_{B7}^C	\hat{Y}_{B8}^C

범주별 복합추정값은 다음과 같이 산출한다.

$$\hat{Y}_{Ai}^C = \hat{w}_{Ai} \hat{Y}_{Ai} + (1 - \hat{w}_{Ai}) \hat{Y}_{Ai}^S, \quad i = 1, 2, \dots, 8,$$

$$\hat{Y}_{Bi}^C = \hat{w}_{Bi} \hat{Y}_{Bi} + (1 - \hat{w}_{Bi}) \hat{Y}_{Bi}^S$$

(Step5) 복합추정값의 Bias, 분산, 평균제곱오차와 CV값을 계산한다.

(1) (Step3)에서 생성된 반복 합성추정값들을 이용하여 반복 복합추정값들을 계산한다.

대영역	세부범주	합성추정값	대영역	세부범주	합성추정값
A	남자15-29세	$\hat{Y}_{A1}^C(l)$	B	남자15-29세	$\hat{Y}_{B1}^C(l)$
	남자30-39세	$\hat{Y}_{A2}^C(l)$		남자30-39세	$\hat{Y}_{B2}^C(l)$
	남자40-49세	$\hat{Y}_{A3}^C(l)$		남자40-49세	$\hat{Y}_{B3}^C(l)$
	남자50세+	$\hat{Y}_{A4}^C(l)$		남자50세+	$\hat{Y}_{B4}^C(l)$
	여자15-29세	$\hat{Y}_{A5}^C(l)$		여자15-29세	$\hat{Y}_{B5}^C(l)$
	여자30-39세	$\hat{Y}_{A6}^C(l)$		여자30-39세	$\hat{Y}_{B6}^C(l)$
	여자40-49세	$\hat{Y}_{A7}^C(l)$		여자40-49세	$\hat{Y}_{B7}^C(l)$
	여자50+	$\hat{Y}_{A8}^C(l)$		여자50+	$\hat{Y}_{B8}^C(l)$

여기에서 l 번째 “성별-연령대별” 반복 복합추정값들은 다음 식으로 계산한다.

$$\hat{Y}_{Ai}^C(l) = \hat{w}_{Ai} \hat{Y}_{Ai} + (1 - \hat{w}_{Ai}) \hat{Y}_{Ai}^S(l), \quad l = 1, 2, \dots, n_i,$$

$$\hat{Y}_{Bi}^C(l) = \hat{w}_{Bi} \hat{Y}_{Bi} + (1 - \hat{w}_{Bi}) \hat{Y}_{Bi}^S(l)$$

(2) i 번째 “성별-연령대별” 범주에서 반복 생성된 복합추정값들을 이용하여 잭나이프 평균제곱오차와 CV값들을 계산한다. 대영역A의 범주별 추정공식은 다음과 같다.

$$mse_J(\hat{Y}_{Ai}^C) = \widehat{Var}_J(\hat{Y}_{Ai}^C) + [\widehat{Bias}_J(\hat{Y}_{Ai}^C)]^2, \quad i = 1, 2, \dots, 8,$$

여기에서

$$\widehat{Var}_J(\hat{Y}_{Ai}^C) = \frac{n_i - 1}{n_i} \sum_{l=1}^{n_i} [\hat{Y}_{Ai}^C(l) - \hat{Y}_{Ai}^C]^2,$$

$$\widehat{Bias}_J(\hat{Y}_{Ai}^C) = (n_i - 1) [\hat{Y}_{Ai}^C(\cdot) - \hat{Y}_{Ai}^C],$$

$$\hat{Y}_{Ai}^C(\cdot) = (\hat{Y}_{Ai}^C(1) + \hat{Y}_{Ai}^C(2) + \dots + \hat{Y}_{Ai}^C(n_i)) / n_i$$

로 주어진다. CV값은

$$CV(\hat{Y}_{Ai}^C) = \frac{\sqrt{mse_J(\hat{Y}_{Ai}^C)}}{\hat{Y}_{Ai}^C} \times 100$$

로 계산한다.

(Step6) 대영역의 “성별-교육정도별” 범주에 대한 합성 및 복합추정값들의 산정은 (Step2)~(Step5)의 추정절차에서 “Borrow Strength”를 “성별-연령대별” 추정값으로 대체하여 산정한다. 추정절차는 (Step2)~(Step5)의 절차와 동일하다.

6.5 성별-연령대별 범주 추정결과

16개 대영역들 중 서울특별시의 “성별-연령대별” 추정결과를 예로 설명한다. 2002년 10월 경제활동인구조사 자료에서 서울특별시의 “성별-연령대별” 범주에 대한 실업자 총계 추정결과가 다음 <표6.3>에 주어졌다.

<표6.3> 서울특별시의 “성별-연령대별” 실업자 총계 추정결과

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	15-29세	35841	5389	15.0	43148	2930	6.3	41551	2261	5.1
	30-39세	29469	4940	16.8	24572	2553	9.6	25555	2015	7.4
	40-49세	14336	3264	22.8	14977	1429	8.8	14948	1199	7.5
	50세이상	6372	2279	35.8	7710	1158	13.9	7460	920	11.6
여	15-29세	28619	4065	14.2	24465	2105	7.9	25303	1660	6.2
	30-39세	10055	2530	25.2	7632	1186	14.3	8065	972	11.3
	40-49세	3867	1736	44.9	6622	867	12.1	6116	693	10.6
	50세이상	3867	1729	44.7	3300	1028	28.8	3428	759	20.8
계	132,426			132,426			132,426			

편의상 “성별-연령대별” 범주의 실업자 총계 추정값들에 대한 CV값의 최대허용한계를 25%로 설정해 놓고 설명하기로 한다.

직접추정법을 적용하여 “성별-연령대별” 범주의 실업자 총계를 추정할 경우에는 전체 8개 범주 중 4개 범주에 대한 추정 결과가 CV값의 목표 기준치인 25%를 만족하는 것으로 나타났다.

합성추정법을 적용한 추정결과에서는 전체 8개 범주 중 7개 범주의 추정결과가 CV값의 기준치를 만족하며, 복합추정법을 적용할 경우에는 모든 범주에 대한 추정결과가 CV값의 기준치를 만족하는 매우 안정적인 추정결과를 나타냈다.

서울특별시의 “성별-연령대별” 범주의 실업자 총계 추정결과, 직접추정결과에서는 8개 범주 중 “남자 50세 이상”, “여자 30-39세”, “여자

40-49세”와 “여자 50세 이상”의 4개 범주의 추정값들이 신뢰할 수 없을 정도의 큰 추정오차를 동반한 반면 합성추정값과 복합추정값들은 매우 안정적인 추정결과를 나타냈다.

이러한 추정결과를 기반으로 직접추정값 대비 합성 또는 복합추정값들의 상대효율이득을 비교하면 <표6.4>와 같다. 여기에서 상대효율이득은 다음 식으로 주어진다.

$$\text{상대효율이득} = \frac{\text{직접추정값의 CV값} - \text{합성 or 복합추정값의 CV값}}{\text{직접추정값의 CV값}} * 100 (\%)$$

<표6.4> 직접추정 대비 합성 및 복합추정의 상대효율이득(서울시)

구 분	직접추정		합성추정			복합추정			
	추정값	CV (%)	추정값	CV (%)	효율 이득(%)	추정값	CV (%)	효율 이득(%)	
남	15-29세	35841	15.0	43148	6.3	58.0	41551	5.1	66.0
	30-39세	29469	16.8	24572	9.6	42.9	25555	7.4	56.0
	40-49세	14336	22.8	14977	8.8	61.4	14948	7.5	67.1
	50세이상	6372	35.8	7710	13.9	61.2	7460	11.6	67.6
여	15-29세	28619	14.2	24465	7.9	44.4	25303	6.2	56.3
	30-39세	10055	25.2	7632	14.3	43.3	8065	11.3	55.2
	40-49세	3867	44.9	6622	12.1	73.1	6116	10.6	76.4
	50세이상	3867	44.7	3300	28.8	35.6	3428	20.8	53.5
계	132,426		132,426		평균52.49	132,426		평균62.26	

<표6.4>의 “성별-연령대별” 8개 범주에 대한 추정결과를 살펴보면 직접추정법보다는 합성 및 복합추정법에 의한 범주별 추정결과가 추정효율이 더 좋은 것으로 나타난다. 직접추정 대비 합성추정의 상대효율이득 양은 평균 52.49%이고, 직접추정 대비 복합추정의 상대효율이득 양은 평균 62.26%이다. 상대효율이득 면에서 볼 때 복합추정법이 합성추정법보다 약 9.77% 정도 추정 효율이 더 좋은 것으로 나타났다.

나머지 15개 대영역들의 추정결과에서도 복합추정의 추정효율이 가장 좋게 나타나며, 이들 대영역들에 대한 “성별-연령대별” 범주의 추정결과

는 다음과 같다.

① 부산광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	15115	2755	18.2	14898	565	3.8	14910	542	3.7
	30-39세	8861	2303	26.0	8773	363	4.2	8777	355	4.1
	40-49세	6776	2202	32.5	6624	381	5.8	6629	370	5.6
	50세이상	3649	1272	34.9	3662	256	7.0	3662	246	6.7
여	15-29세	8881	2087	23.5	8333	268	3.2	8343	263	3.2
	30-39세	1974	994	50.4	2734	171	6.3	2712	166	6.2
	40-49세	2467	977	39.6	2952	181	6.2	2936	175	6.0
	50세이상	1480	1098	74.2	1227	191	15.7	1234	185	15.1
계		49,203			49,203			49,203		

② 인천광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	14399	2906	20.2	13801	396	4.1	13832	389	4.0
	30-39세	9729	1929	19.8	10449	275	3.7	10447	269	3.6
	40-49세	7783	1810	23.3	6857	235	4.9	6882	231	4.7
	50세이상	4281	1539	36.0	2894	157	7.7	2908	155	7.5
여	15-29세	7168	1588	22.2	7639	152	2.8	7613	151	2.8
	30-39세	2263	944	41.7	2981	91	4.3	2963	90	4.3
	40-49세	1886	764	40.5	2772	108	5.5	2752	105	5.4
	50세이상	755	542	71.9	871	52	8.6	867	52	8.5
계		48,264			48,264			48,264		

③ 대구광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	15-29세	9133	2234	24.5	9103	286	2.8	9120	281	2.7
	30-39세	2740	1119	40.9	5764	167	2.6	5712	163	2.5
	40-49세	6393	1764	27.6	3921	181	4.1	3953	179	4.0
	50세이상	1827	920	50.4	1939	112	5.1	1941	110	5.1
여	15-29세	6744	1742	25.8	5197	187	3.2	5225	185	3.1
	30-39세	899	654	72.8	1818	143	7.0	1777	137	6.8
	40-49세	2248	1010	44.9	1622	124	6.8	1634	122	6.7
	50세이상	0	0	-	621	0	-	623	0	-
계	29,984			29,984			29,984			

④ 광주광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	15-29세	5815	1591	27.4	4410	107	1.7	4416	106	1.7
	30-39세	1385	619	44.7	2527	52	1.5	2520	51	1.5
	40-49세	1108	556	50.2	1520	36	1.7	1519	36	1.7
	50세이상	1108	676	61.1	745	37	3.5	746	37	3.5
여	15-29세	2464	829	33.7	2613	61	1.7	2612	61	1.7
	30-39세	274	274	100.2	790	11	1.1	790	11	1.1
	40-49세	821	475	58.0	660	27	2.9	661	27	2.9
	50세이상	548	387	70.8	256	20	5.6	256	20	5.6
계	13,523			13,523			13,523			

⑤ 대전광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	4406	1219	27.7	6204	102	1.6	6194	102	1.6
	30-39세	2203	968	44.0	3600	62	1.7	3596	62	1.7
	40-49세	2203	742	33.7	2389	53	2.2	2389	53	2.2
	50세이상	315	222	70.8	1086	19	1.8	1081	19	1.7
여	15-29세	5764	1281	22.2	3472	76	2.2	3482	76	2.1
	30-39세	1729	696	40.3	1148	34	2.9	1150	34	2.9
	40-49세	2594	901	34.8	969	58	5.9	976	58	5.9
	50세이상	0	0	-	346	0	-	347	0	-
계		19,214			19,214			19,214		

⑥ 울산광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	2903	851	29.3	3480	55	1.4	3478	55	1.4
	30-39세	581	415	71.5	2758	26	0.8	2750	26	0.8
	40-49세	2032	746	36.7	1781	44	2.2	1782	44	2.2
	50세이상	871	467	53.6	681	16	2.1	681	16	2.1
여	15-29세	1396	749	53.7	1891	31	1.4	1891	30	1.4
	30-39세	2234	844	37.8	783	45	5.1	787	45	5.1
	40-49세	2234	708	31.7	710	42	5.2	715	42	5.2
	50세이상	0	0	-	167	0	-	167	0	-
계		12,251			12,251			12,251		

⑦ 경기도

구분	직접추정			합성추정			복합추정			
	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	
남	15-29세	36848	5763	15.6	42265	3910	8.7	40739	2677	6.3
	30-39세	22630	4248	18.8	21774	2586	11.3	22114	1886	8.2
	40-49세	13803	3627	26.3	12508	1857	14.1	12877	1471	11.0
	50세이상	11137	3540	31.8	7462	1486	18.9	8082	1263	15.1
여	15-29세	19607	4977	25.4	18039	3062	16.1	18547	2221	11.6
	30-39세	4678	2041	43.6	7105	1105	14.7	6628	854	12.5
	40-49세	7284	2996	41.1	6583	1481	21.3	6778	1190	17.0
	50세이상	1711	1124	65.7	1961	551	26.6	1933	444	22.2
계	117,698			117,698			117,698			

⑧ 강원도

구분	직접추정			합성추정			복합추정			
	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	
남	15-29세	4902	1357	27.7	5366	398	6.7	5322	367	6.2
	30-39세	2810	969	34.5	2688	201	6.7	2697	193	6.5
	40-49세	1690	775	45.9	1618	161	9.0	1623	155	8.6
	50세이상	887	538	60.7	971	90	8.4	972	88	8.2
여	15-29세	1963	770	39.2	1512	173	10.3	1534	165	9.7
	30-39세	381	382	100.4	341	80	21.2	343	77	20.3
	40-49세	381	384	100.9	698	36	4.7	699	35	4.6
	50세이상	381	383	100.6	201	51	22.9	205	50	22.2
계	13,395			13,395			13,395			

⑨ 충청북도

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	2216	864	39.0	4054	262	4.2	3944	240	4.0
	30-39세	1035	595	57.5	1973	137	4.5	1963	130	4.4
	40-49세	624	453	72.8	1109	108	6.3	1102	102	6.2
	50세이상	2216	850	38.4	622	144	14.9	664	140	13.9
여	15-29세	2210	680	30.8	1185	202	11.0	1240	186	9.9
	30-39세	0	0	-	250	0	-	257	0	-
	40-49세	606	368	60.8	461	108	15.1	469	100	14.1
	50세이상	880	520	59.1	131	88	43.3	148	86	38.5
계		9,787			9,787			9,787		

⑩ 충청남도

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	3815	941	24.7	6534	190	2.4	6450	182	2.3
	30-39세	4498	1108	24.7	2969	167	4.6	3009	163	4.5
	40-49세	1251	747	59.7	1862	90	4.0	1865	89	3.9
	50세이상	1216	612	50.3	1501	109	6.0	1497	105	5.8
여	15-29세	3880	977	25.2	2442	106	3.6	2470	105	3.5
	30-39세	1358	612	45.1	813	51	5.2	821	50	5.1
	40-49세	941	668	71.0	896	53	4.9	902	52	4.8
	50세이상	444	278	62.6	387	29	6.2	390	29	6.1
계		17,403			17,403			17,403		

⑪ 전라북도

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	5929	1887	31.8	7702	698	8.9	7539	614	8.1
	30-39세	4344	1260	29.0	3499	367	10.4	3589	338	9.4
	40-49세	1908	998	52.4	1945	242	12.3	1957	229	11.6
	50세이상	2114	1058	50.1	1366	207	15.0	1403	199	14.1
여	15-29세	1737	823	47.4	2244	267	11.7	2211	241	10.9
	30-39세	955	671	70.3	477	134	27.8	499	129	25.8
	40-49세	1433	803	56.1	885	179	20.0	917	171	18.5
	50세이상	0	0	-	302	0	-	305	0	-
계		18,420			18,420			18,420		

⑫ 전라남도

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	6837	1697	24.8	8626	780	11.0	8450	644	9.1
	30-39세	5772	1588	27.5	4350	428	12.0	4467	399	10.7
	40-49세	3537	1143	32.3	2494	358	17.5	1618	326	15.0
	50세이상	780	548	70.3	1897	200	12.9	1759	176	12.1
여	15-29세	2634	862	32.7	2213	253	13.9	2259	233	12.4
	30-39세	888	634	71.5	485	134	33.7	503	128	30.7
	40-49세	1142	670	58.7	1101	230	25.5	1115	206	22.2
	50세이상	0	0	-	425	0	-	419	0	-
계		21,590			21,590			21,590		

⑬ 경상북도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	15-29세	6995	1889	27.0	9835	442	3.7	9708	419	3.6
	30-39세	4595	1453	31.6	4344	227	4.3	4369	222	4.3
	40-49세	3520	1326	37.7	2825	198	5.8	2851	194	5.7
	50세이상	2533	1144	45.2	2229	218	8.1	2244	210	7.8
여	15-29세	5029	1286	25.6	3786	228	5.0	3831	221	4.8
	30-39세	1823	910	49.9	1284	130	8.4	1300	127	8.2
	40-49세	1362	788	57.9	1430	135	7.9	1434	131	7.7
	50세이상	461	463	100.6	587	97	13.7	583	92	13.3
계	26,318			26,318			26,318			

⑭ 경상남도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	15-29세	9459	2734	28.9	14359	593	4.6	14173	567	4.4
	30-39세	7669	2873	37.5	6942	378	6.0	6963	372	5.9
	40-49세	6577	2960	45.0	4496	483	11.9	4565	470	11.4
	50세이상	2917	1144	39.2	2979	203	7.5	2984	197	7.3
여	15-29세	3472	1504	43.3	5721	232	4.5	5673	227	4.4
	30-39세	3824	1590	41.6	2200	296	14.9	2266	286	13.9
	40-49세	4518	1787	39.6	2316	281	13.4	2379	274	12.7
	50세이상	1389	981	70.7	812	108	14.7	821	107	14.4
계	39,825			39,825			39,825			

⑮ 제주도

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	15-29세	1215	526	43.4	811	15	0.6	811	15	0.6
	30-39세	0	0	-	380	0	-	380	0	-
	40-49세	486	342	70.5	217	9	1.4	217	9	1.4
	50세이상	0	0	-	164	0	-	164	0	-
여	15-29세	0	0	-	329	0	-	329	0	-
	30-39세	237	245	103.5	118	7	2.0	118	7	2.0
	40-49세	237	242	102.4	113	5	1.7	113	5	1.7
	50세이상	0	0	-	43	0	-	43	0	-
계		2,175			2,175			2,175		

6.6 성별-교육정도별 범주 추정결과

2002년 10월 경제활동인구조사 자료에서 서울특별시의 “성별-교육정도별” 범주에 대한 실업자 총계 추정결과를 살펴보면 <표6.5>와 같다.

<표6.5> 서울특별시의 “성별-교육정도별” 실업자 총계 추정결과

구분	직접추정			합성추정			복합추정			
	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	
남	초졸이하	4779	2268	47.5	5013	987	18.2	5151	829	15.6
	중졸	6372	2553	40.0	7976	1211	14.0	7952	988	12.0
	고졸	37433	5639	15.1	53578	3784	6.5	49931	2609	5.0
	대졸이상	37433	5488	14.7	23839	1934	7.5	26210	1720	6.3
여	초졸이하	773	780	100.9	1140	323	26.1	1129	275	23.6
	중졸	1547	1104	71.4	4392	558	11.7	3968	444	10.8
	고졸	24751	4495	18.2	25680	2757	9.9	26085	2003	7.4
	대졸이상	19337	3966	20.5	10807	1263	10.8	11999	1146	9.2
계	132,425			132,425			132,425			

직접추정법을 적용하여 “성별-교육정도” 범주의 실업자 총계를 추정할 경우, 전체 8개 범주 중 “남자 초등졸 이하”, “남자 중졸”, “여자 초등졸 이하”와 “여자 중졸”의 4개 범주에 대한 추정 결과가 CV값의 목표기준치인 25%를 만족하지 못하는 매우 불안정한 추정결과를 나타낸다.

합성추정법을 적용한 추정결과에서는 전체 8개 범주 중 “여자 초등졸 이하”의 1개 범주만이 CV값의 목표기준치를 벗어나고, 나머지 7개 범주의 추정결과는 안정적이다. 복합추정법을 적용할 경우에는 모든 범주에 대한 추정결과가 CV값의 목표기준치를 만족하는 매우 안정적인 추정결과를 나타낸다.

서울특별시의 “성별-교육정도별” 범주의 실업자 총계 추정결과, 직접추정결과에서는 8개 범주 중 4개 범주의 추정값들이 신뢰할 수 없을 정도의 큰 추정오차를 동반한 반면 합성추정값과 복합추정값들은 매우 안정적

인 추정결과를 나타냈다.

이러한 추정결과를 기반으로 직접추정값 대비 합성 또는 복합추정값들의 상대효율이익을 비교하면 <표6.6>과 같다. 여기에서 상대효율이익은 다음 식과 같이 주어진다.

$$\text{상대효율이익} = \frac{\text{직접추정값의 CV값} - \text{합성 or 복합추정값의 CV값}}{\text{직접추정값의 CV값}} * 100 (\%)$$

<표6.6> 직접추정 대비 합성 및 복합추정의 상대효율이익(서울시)

구 분	직접추정		합성추정			복합추정			
	추정값	CV (%)	추정값	CV (%)	효율 이익(%)	추정값	CV (%)	효율 이익(%)	
남	초졸이하	4779	47.5	5013	18.2	61.7	5151	15.6	67.2
	중졸	6372	40.0	7976	14.0	65.0	7952	12.0	70.0
	고졸	37433	15.1	53578	6.5	57.0	49931	5.0	66.9
	대졸이상	37433	14.7	23839	7.5	49.0	26210	6.3	57.1
여	초졸이하	773	100.9	1140	26.1	74.1	1129	23.6	76.6
	중졸	1547	71.4	4392	11.7	83.6	3968	10.8	84.9
	고졸	24751	18.2	25680	9.9	45.6	26085	7.4	59.3
	대졸이상	19337	20.5	10807	10.8	47.3	11999	9.2	55.1
계	132,425		132,425		평균60.41	132,425		평균67.14	

제 6.5절의 “성별-연령대별” 추정결과와 마찬가지로 “성별-교육정도별” 추정결과에서도 복합추정법이 다른 추정법들에 비해 매우 안정적인 것으로 나타났다.

직접추정 대비 합성추정의 상대효율이익 양은 평균 60.41%이고, 직접추정 대비 복합추정의 상대효율이익 양은 평균 67.14%이다. 상대효율이익 면에서 볼 때 복합추정법이 합성추정법보다 약 6.73% 정도 추정 효율이 더 좋은 것으로 나타났다.

나머지 15개 대영역들의 추정결과에서도 복합추정의 추정효율이 가장 좋게 나타나며, 이들 대영역들에 대한 “성별-교육정도별” 범주의 추정결과는 다음과 같다.

① 부산광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	3127	1455	46.5	2645	237	9.0	2673	230	8.7
	중졸	2606	1387	53.2	4174	276	6.6	4138	265	6.5
	고졸	19285	3153	16.4	21437	702	3.3	21461	669	3.1
	대졸이상	9382	2136	22.8	5701	149	2.6	5752	148	2.6
여	초졸이하	493	490	99.5	648	97	15.1	646	94	14.7
	중졸	2467	1323	53.6	2275	259	11.4	2296	249	11.0
	고졸	4934	1475	29.9	9598	288	3.0	9481	277	3.0
	대졸이상	6908	1851	26.8	2725	114	4.2	2756	114	4.2
계	49,202			49,202			49,202			

② 인천광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	2335	902	38.7	2589	140	7.7	2597	136	7.4
	중졸	3502	1094	31.2	3807	173	6.5	3823	169	6.2
	고졸	19847	3324	16.8	22447	485	3.1	22491	475	3.0
	대졸이상	10507	1887	18.0	5158	86	2.4	5161	86	2.4
여	초졸이하	755	541	71.7	563	78	19.8	571	77	19.0
	중졸	755	532	70.6	2018	67	4.7	1996	66	4.7
	고졸	5659	1500	26.5	9786	182	2.6	9728	179	2.6
	대졸이상	4904	1218	24.8	1896	41	3.1	1895	40	3.0
계	48,264			48,264			48,264			

③ 대구광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	1827	913	50.0	1712	150	7.7	1718	146	7.6
	중졸	3197	1409	44.1	2391	182	6.7	2410	179	6.6
	고졸	8220	2016	24.5	12740	314	2.2	12671	307	2.1
	대졸이상	6850	1728	25.2	3883	92	2.1	3908	92	2.1
여	초졸이하	0	0	-	400	0	-	402	0	-
	중졸	3597	1396	38.8	1326	201	13.4	1370	197	12.8
	고졸	3147	1279	40.6	5646	193	3.0	5609	189	3.0
	대졸이상	3147	1073	34.1	1887	47	2.2	1898	47	2.2
계	29,985			29,985			29,985			

④ 광주광역시

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	554	554	100.1	601	31	3.7	601	31	3.7
	중졸	1108	535	48.3	830	33	2.8	832	33	2.8
	고졸	4153	1070	25.8	5523	94	1.2	5516	93	1.2
	대졸이상	3600	1246	34.6	2248	52	1.7	2254	52	1.7
여	초졸이하	548	392	71.5	143	22	11.2	145	22	11.1
	중졸	548	390	71.1	483	20	3.0	484	20	2.9
	고졸	1642	612	37.3	2684	57	1.5	2678	56	1.5
	대졸이상	1369	559	40.8	1008	22	1.6	1011	22	1.6
계	13,522			13,522			13,522			

⑤ 대전광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	초졸이하	629	407	64.7	1035	37	3.5	1032	37	3.5
	중졸	944	450	47.7	1274	37	2.9	1272	37	2.9
	고졸	4406	1626	36.9	7601	120	1.6	7588	119	1.5
	대졸이상	3147	935	29.7	3370	55	1.6	3371	55	1.6
여	초졸이하	0	0	-	221	0	-	221	0	-
	중졸	865	643	74.5	685	41	5.9	686	41	5.9
	고졸	5188	1344	25.9	3686	100	2.7	3697	99	2.6
	대졸이상	4035	1017	25.2	1344	32	2.3	1348	32	2.3
계		19,214			19,214			19,214		

⑥ 울산광역시

구 분		직접추정			합성추정			복합추정		
		추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)
남	초졸이하	0	0	-	618	0	-	618	0	-
	중졸	2032	1037	51.1	945	57	5.3	947	57	5.3
	고졸	3484	1015	29.2	5862	62	0.9	5853	61	0.9
	대졸이상	871	421	48.4	1275	9	0.6	1276	9	0.6
여	초졸이하	558	400	71.7	141	20	12.9	142	20	12.8
	중졸	1117	517	46.3	530	31	5.2	531	31	5.1
	고졸	3351	850	25.4	2352	51	1.9	2355	50	1.9
	대졸이상	838	626	74.8	528	8	1.3	529	8	1.3
계		12,251			12,251			12,251		

⑦ 경기도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	1067	628	58.9	3439	438	12.1	2752	294	10.6
	중졸	8767	2649	30.2	6282	1309	19.7	6955	1052	14.9
	고졸	48636	6898	14.2	58435	4817	7.8	56632	3238	5.6
	대졸이상	25947	4734	18.2	15852	1998	11.9	17882	1696	9.4
여	초졸이하	1711	1106	64.7	1354	475	33.3	1453	401	27.2
	중졸	2967	1738	58.6	5095	835	15.5	4859	678	13.8
	고졸	17989	4923	27.4	22403	3343	14.1	21567	2288	10.5
	대졸이상	10612	3418	32.2	4837	1156	22.6	5595	1037	18.3
계	117,696			117,696			117,696			

⑧ 강원도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	1130	583	51.6	755	121	14.4	777	116	13.6
	중졸	803	567	70.7	1215	135	10.0	1206	128	9.7
	고졸	5546	1601	28.9	6977	452	5.8	6931	418	5.5
	대졸이상	2811	942	33.6	1695	135	7.2	1738	133	7.0
여	초졸이하	381	384	100.8	630	51	7.3	634	50	7.2
	중졸	0	0	-	74	0	-	75	0	-
	고졸	981	601	61.2	1536	181	10.6	1503	166	10.1
	대졸이상	1744	730	41.9	514	73	12.8	532	72	12.5
계	13,396			13,396			13,396			

⑨ 충청북도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	968	581	60.1	527	102	12.5	546	99	12.0
	중졸	690	492	71.4	792	85	6.9	805	82	6.8
	고졸	3397	1076	31.7	5195	315	3.9	5102	290	3.8
	대졸이상	1035	505	48.9	1243	104	5.4	1256	99	5.3
여	초졸이하	1213	577	47.6	417	134	20.6	449	127	18.7
	중졸	0	0	-	54	0	-	56	0	-
	고졸	1213	613	50.6	1181	192	10.5	1178	175	9.8
	대졸이상	1271	454	35.7	375	62	10.6	395	60	10.2
계	9,787			9,787			9,787			

⑩ 충청남도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	1488	781	52.5	1374	189	11.3	1372	178	10.7
	중졸	1759	652	37.1	1250	94	6.2	1260	92	6.0
	고졸	5373	1537	28.6	8689	210	2.0	8654	206	2.0
	대졸이상	2161	744	34.4	1554	39	2.1	1562	39	2.1
여	초졸이하	222	224	101.2	451	37	6.8	446	36	6.7
	중졸	941	670	71.2	856	53	5.1	860	52	5.1
	고졸	3658	986	27.0	2853	115	3.3	2870	114	3.3
	대졸이상	1802	770	42.8	378	24	5.4	381	24	5.4
계	17,404			17,404			17,404			

⑪ 전라북도

구분		직접추정			합성추정			복합추정		
		추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)
남	초졸이하	528	536	101.4	992	94	9.4	986	91	9.2
	중졸	1585	920	58.0	1532	190	12.2	1545	182	11.7
	고졸	8366	2135	25.5	9442	730	7.6	9394	654	6.9
	대졸이상	3816	1275	33.4	2547	334	13.0	2647	313	11.8
여	초졸이하	304	308	101.5	801	161	19.8	699	126	18.0
	중졸	0	0	-	106	0	-	107	0	-
	고졸	1911	985	51.6	2080	272	12.9	2083	252	12.1
	대졸이상	1911	871	45.6	922	158	16.9	961	153	15.9
계		18,421			18,421			18,421		

⑫ 전라남도

구분		직접추정			합성추정			복합추정		
		추정값	추정오차	CV (%)	추정값	추정오차	CV (%)	추정값	추정오차	CV (%)
남	초졸이하	1274	737	57.9	1558	280	21.9	1553	244	19.1
	중졸	2757	1249	45.3	2191	452	25.1	2320	400	20.9
	고졸	8501	1768	20.8	11445	797	8.5	11228	662	7.2
	대졸이상	4393	1357	30.9	2171	187	10.5	2224	183	10.0
여	초졸이하	793	572	72.2	1222	218	21.8	1186	190	19.5
	중졸	444	446	100.7	132	109	100.7	155	102	80.9
	고졸	2285	808	35.4	2206	276	15.3	2260	247	13.3
	대졸이상	1142	585	51.2	662	41	7.7	664	41	7.6
계		21,589			21,589			21,589		

⑬ 경상북도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	2489	855	34.4	1620	249	12.7	1673	229	11.5
	중졸	986	708	71.9	1803	111	5.1	1799	109	5.1
	고졸	10043	2244	22.3	13439	507	3.1	13333	483	3.0
	대졸이상	4125	1331	32.3	2370	91	3.2	2400	91	3.2
여	초졸이하	922	656	71.2	627	137	18.1	640	131	17.2
	중졸	1362	794	58.3	1335	124	7.7	1345	121	7.5
	고졸	3646	1279	35.1	4467	248	4.6	4462	239	4.5
	대졸이상	2745	1080	39.4	658	38	4.9	668	38	4.8
계	26,318			26,318			26,318			

⑭ 경상남도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	2545	1167	45.9	1809	229	14.0	1847	220	13.2
	중졸	1824	1063	58.3	2657	205	8.5	2636	198	8.3
	고졸	14583	3558	24.4	20242	765	4.2	20069	731	4.0
	대졸이상	7669	2995	39.1	4068	208	5.7	4091	207	5.6
여	초졸이하	0	0	-	754	0	-	755	0	-
	중졸	5907	2149	36.4	2049	355	19.1	2169	345	17.6
	고졸	6601	1821	27.6	7031	335	5.3	7043	324	5.1
	대졸이상	694	704	101.5	1215	38	3.5	1214	38	3.5
계	39,823			39,823			39,823			

⑮ 제주도

구 분	직접추정			합성추정			복합추정			
	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	추정값	추정 오차	CV (%)	
남	초졸이하	0	0	-	79	0	-	79	0	-
	중졸	243	242	100.0	132	6	1.6	132	6	1.6
	고졸	729	418	57.5	1119	16	0.5	1119	16	0.5
	대졸이상	729	420	57.8	242	4	0.6	242	4	0.6
여	초졸이하	237	240	101.7	39	5	4.8	39	5	4.8
	중졸	0	0	-	99	0	-	99	0	-
	고졸	237	236	99.9	389	7	0.6	389	7	0.6
	대졸이상	0	0	-	74	0	-	75	0	-
계	2,175			2,175			2,175			

제7장 결 언

지식정보화 및 글로벌 정보화 시대에 대비하고 지방자치제도의 정착으로 다양하고 세분화된 통계 작성 요구에 대처하기 위해서 먼저 2003년에 신규로 도입되는 경제활동인구조사 체계를 요약하면서 우리나라 통계 조직과 경제적인 측면을 고려하여 소지역 추정법 적용의 필요성을 언급하였다. 특히 미국, 캐나다, 프랑스와 일본 등 통계 선진국의 소지역 추정법 적용 실태를 상세하게 요약 정리하여 우리나라에서 소지역 추정법의 활용 시 시행착오를 최소화하도록 하였고, 기존의 소지역 추정법에 관한 연구를 총망라하여 요약 정리함으로써 소지역 추정법의 적용분야를 확대하는데 도움을 주고자 하였다. 특히 본 연구에서 적용되지 않은 모형기반 간접 추정법에 대한 내용을 상세하게 언급하여 앞으로 있을 유관 분야 연구의 지침이 되도록 하였다.

본 연구의 핵심내용인 시군구 단위의 경제활동인구 관련 통계작성에서는 복합추정법의 적용을 다루었으며 시군구 단위의 경제활동인구 관련 통계 작성 시 “Borrow Strength”의 효과를 극대화하기 위해 전국의 시군구 단위 소지역들을 유사 집락들로 집락화하는 방안을 별도로 연구하였으나 실용성에 문제가 있어서 특·광역시 및 도 단위별로 시군구 단위 소지역들을 집락화하는 방안을 모색하였다. SAS의 FASTCLUS Procedure를 이용하여 특·광역시 및 도 단위별로 시군구의 집락화를 시도하였으며 결과적으로 합성추정값의 편향을 줄여 복합추정값의 안정성과 정확성을 제고하였다. 인구 10만 이상의 시군구 단위 소지역들에 대해 매월 또는 분기별로 노동력통계를 생산하고 공표할 것을 염두에 두고 추정절차와 알고리즘을 개발하였으며 이를 구현하기 위해서 SAS Macro를 이용하여 프로그램을 작성하였다. 특히 본 연구의 특징적인 연구 내용 중 하나는 합성추정값과 복합추정값의 평균제곱오차 추정 시 과소추정의 문제를 해결하기 위해서 잭나이프 추정법을 적용하는 방안을 연구하였다는 점이며 이에 대한 프로그램도 개발하여 실용화에 기여하고자 하였다.

본 연구의 타당성과 적용 가능성을 입증하기 위해서 2002년10월에 통계청에서 실시한 신규 표본가구의 병행조사 자료와 2000년 「인구주택 총 조

사」의 10%표본조사 자료를 이용하여 16개 특·광역시와 도별로 각 시군구 단위 소지역들의 실업자 총계에 대한 추정값의 계산 결과를 제시하였다. 다음으로는 특·광역시와 도 단위의 “성별-연령대별”과 “성별-교육정도별” 범주의 경제활동인구 관련 통계 작성을 위해 16개 특·광역시와 도 단위들을 실업률이 유사한 동질적인 집락으로 집락화하는 방안을 연구하였다. 여기에서도 SAS의 FASTCLUS Procedure를 이용하여 광역시와 도 단위들을 실업률이 유사한 동질적인 집락으로 묶어 합성 추정값의 편향을 줄이고자 하였다. 연구결과의 실현가능성과 타당성을 입증하기 위해 2000년 센서스 자료와 2002년10월 병행조사 자료 및 상주추계인구(통계청 인구분석과 제공)를 이용하여 실업자 총계를 추정하여 수치적으로 비교하였다. 취업자 총계와 비경제활동인구 등의 추정절차는 실업자 총계 추정절차와 동일하므로 세부적인 추정절차와 알고리즘은 보고서에 포함시키지 않았고 SAS 프로그램과 추정결과를 부록에 실어 참고토록 하였다. 앞으로 통계청에서 소지역추정법을 이용하여 관심분야의 통계 작성 연구에 참고로 활용할 수 있도록 시군구 단위 경제활동인구 관련 통계 작성과 특·광역시와 도 단위의 “성별-연령대별”, “성별-교육정도별” 노동력 통계작성에 관한 SAS 프로그램의 SOURCE를 부록에 포함하였다.

소지역 추정법의 장점은 조사된 자료 외에 행정업무 또는 보고통계 내용 등을 보조정보로 활용하여 추정값의 편향과 추정오차를 줄이는 것이지만 이 보고서에서는 노동부의 고용정보원에서 관리하는 구직등록 데이터베이스 자료와 실직보험 청구 자료를 보조정보로 이용하는 추정방법에 대한 연구는 다루지 못하였으므로 별도의 연구과제로 추진하여야 하며 이에 대한 통계 작성 체계는 앞으로 심층적으로 추진되어야 할 연구 과제로 남겨두었다. 미국이나 캐나다 등에서는 실업 관련 통계 작성에서 핵심적인 보조정보로 구직등록 데이터베이스와 실직보험신청 데이터베이스를 이용하고 있으나 우리나라에서는 이들의 데이터베이스에 대한 신뢰성과 작성 주기 등에서 장기적으로 표준화 및 관리체계를 개선한 후에 경제활동인구조사 자료와 연계하는 것이 바람직할 것이다.

본 연구를 시작으로 하여 다양한 분야의 정부통계작성에서도 행정업무자료와 보고통계자료 등을 보조정보로 이용하는 소지역 추정법에 대한 연구가 활성화되어 조사통계의 고 비용성과 조사환경의 열악화로 인한 실사의 어려운 문제점들이 해소되기를 기대하며 정부통계의 글로벌 정보화 시대 편승에 다소 도움이 되기를 기대한다.

참 고 문 헌

- (1) 강순희 외 2인(1999), 실업통계의 개선방안, 한국노동연구원
- (2) 이계오(1998), 지역 통계의 활성화 방안, 한국 통계학회 충청지회, 충북 발연구원 주최 세미나.
- (3) 이계오(1999), “소지역통계” 워크샵, 한국통계학회 조사통계연구회, 충북개발연구원 주최 세미나.
- (4) 이계오(2000), 시군구 실업자 추정을 위한 소지역 추정법, 응용통계연구 제13권 2호, 275-286
- (5) 이계오(2001), 소지역 실업률 추정기법 및 전산프로그램 개발, 2001 노동부 용역연구보고서
- (6) 이계오, 정연수(2000), 소지역 추정법을 이용한 시군구의 실업자 추정, 한국조사연구학회 추계학술대회 발표논문집, 229-250
- (7) 조옥현과 노근호(1999), “소지역통계” 워크샵, 한국통계학회 조사통계연구회, 충북개발연구원 주최 세미나.
- (8) 일본 총무청 통계국(1997), 일본 노동력조사의 표본설계.
- (9) Cronkhite, F. R.(1987), Use of regression techniques for developing state and area employment and unemployment estimates, *In Small Area Statistics*, 160-174
- (10) Decaudin, G. and Labat, Jean-Claude(1997), A synthetic, robust and efficient method of making small area population estimates in France, *Survey Methodology*, Vol. 23, No. 2, 91-98
- (11) Ericksen, E. P.(1974), A regression method for estimating population of local areas, *The Journal of the American Statistical Association*, Vol. 69, 867-875
- (12) 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
- (13) Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B., and Lindeyer J.(1998), *Methodology of the canadian labour force survey*, Statistics Canada, Household Survey Methods Division, Catalogue no. 71-526-XPB.
- (14) Ghosh, M., Natarajan, K., and Stroud, T. W. F., and Carlin, B.P.(1998), Generalized linear models for small area estimation, *The Journal of the American Statistical Association*, Vol. 93, 273-282
- (15) Ghosh, M. and Rao, J. N. K.(1994), Small area estimation: An appraisal, *Statistical science*, Vol. 9, 55-93
- (16) Gonzalez, M. E. and Hoza, C.(1978), Small area estimation with application to unemployment and housing estimates, *The Journal of the American Statistical Association*, Vol. 73, 7-15
- (17) Marker, D. A.(1999), Organization of small area estimators using a generalized linear regression framework, *Journal of Official Statistics*, Vol. 15, No. 1, 1-24
- (18) Moura, F.A.S., and Holt, D.(1999), Small Area Estimation Using Multilevel Models, *Survey Methodology*, Vol. 25, 73-80
- (19) Prasad, N. G. N. and Rao, J. N. K.(1990), The estimation of the mean squared error of small-area estimators, *The Journal of the American Statistical Association*, Vol. 85, 163-171
- (20) Singh, M. P., Gambino, J., and Mantel, H. J.(1994), Issues and strategies for small area data, *Survey Methodology*, Vol. 20, No. 1, 3-22
- (21) Statistics Canada (2001), *Guide to the labour force survey*, Catalogue no. 71-543-GIE
- (22) Tiller, R. B. and Brown, S.(1993), Bureau of Labor Statistics' state and local area estimates of employment and unemployment,

Indirect Estimators in Fedral Programs, 5. 1-19

(23) You, Y. and Rao, J.N.K. (2000), Hierarchical Bayes Estimation of Small Area Means Using Multi-Level Models, *Survey Methodology*, Vol.26, 173-181

(24) Zimmerman, T. S., Evans, T. D., and Tiller, R. B.(1994), State unemployment rate time series, *ASA 1994 Proceeding of the Section on Survey Research Methods*, 1077-1082

(25) U.S. Department of Labour Bureau of Labor Statistics (2000), *Design and methodology*, Technical paper 63.

부록1 시군구 추정결과(전국)

2002년 10월의 경제활동인구조사 자료로부터 대영역 내의 시군구단위 소지역들에 대한 “실업자 수”, “취업자 수”, “경제활동인구”, “비경제활동인구”의 복합추정 결과를 요약하면 다음과 같다.

(1) 서울특별시

시군구	실업자 수		취업자 수		경제활동인구		비경제활동인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
서초구	4054	10.9	162700	1.3	166754	1.1	138649	1.3	2.4
강남구	5694	9.5	228722	1.6	234416	1.5	198713	1.7	2.4
송파구	6481	9.4	272964	1.2	279445	1.2	226926	1.5	2.3
강동구	5114	13.9	203610	0.6	208724	0.4	170009	0.5	2.5
양천구	5721	2.6	210052	1.2	215773	1.2	138833	1.8	2.7
강서구	6370	5.1	230449	0.8	236819	0.8	152116	1.1	2.7
구로구	5567	8.2	191823	0.6	197390	0.5	124064	0.8	2.8
금천구	3391	7.5	121442	0.5	124833	0.4	78106	0.6	2.7
영등포구	5181	5.1	185210	0.8	190391	0.9	119510	1.3	2.7
동작구	5194	3.0	184745	1.0	189939	1.0	123212	1.5	2.7
관악구	7234	6.8	244433	1.0	251667	1.1	162449	1.7	2.9
성동구	5921	4.4	142799	0.6	148720	0.5	99943	0.8	4.0
광진구	6909	6.6	163182	0.7	170091	0.8	116702	1.1	4.1
중랑구	7503	3.0	185182	0.4	192685	0.4	131140	0.6	3.9
성북구	8034	3.7	198504	0.4	206538	0.4	141301	0.5	3.9
강북구	6109	5.3	151103	0.6	157212	0.6	107992	0.9	3.9
도봉구	6041	3.3	151255	0.7	157296	0.7	106384	1.0	3.8
노원구	10654	6.1	262113	0.9	272767	0.9	187177	1.3	3.9
종로구	1749	4.0	79519	0.7	81268	0.7	55028	1.1	2.2
충구	1424	9.4	63875	0.5	65299	0.3	43648	0.4	2.2
용산구	2346	9.4	105630	1.4	107976	1.5	73973	2.1	2.2
은평구	4620	11.4	205149	1.4	209769	1.2	140234	1.7	2.2
서대문구	3534	7.6	161768	0.8	165302	0.8	112735	1.2	2.1
마포구	3841	10.3	167428	0.9	171269	0.8	116123	1.1	2.2
동대문구	3738	7.4	174014	0.9	177752	0.9	119351	1.3	2.1
계	132,424		4,447,671		4,580,095		3,184,318		2.89

(2) 부산광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
중구/기장군	1763	4.6	57229	0.5	58992	0.4	47720	0.5	3.0
중구	756		24551		25308		20472		
기장군	1007		32678		33684		27248		
서구	2081	3.2	65751	0.4	67832	0.3	56065	0.4	3.1
영도구	2500	5.0	77414	0.6	79914	0.6	66361	0.7	3.1
동래구	4052	4.6	124865	0.5	128917	0.5	107388	0.6	3.1
남구	4182	3.1	130407	0.4	134589	0.4	111859	0.4	3.1
사하구	5077	4.1	158332	0.9	163409	0.8	133949	1.0	3.1
금정구	3955	7.1	121163	0.6	125118	0.6	104664	0.7	3.2
연제구	3137	8.9	97327	0.5	100464	0.5	82769	0.6	3.1
수영구	2433	2.2	76853	0.5	79286	0.4	65622	0.5	3.1
동구	1602	4.0	54733	0.4	56335	0.5	45522	0.6	2.8
부산진구	5468	10.3	182822	0.8	188290	0.8	152082	1.0	2.9
북구	3769	5.7	127999	0.9	131768	0.7	104927	0.9	2.9
해운대구	4747	7.2	167191	0.9	171938	0.9	137898	1.1	2.8
강서구	689	1.7	23834	0.3	24523	0.2	19495	0.3	2.8
사상구	3748	2.8	125543	0.7	129291	0.7	103266	0.9	2.9
계	49,203		1,591,463		1,640,666		1,339,587		3.00

(3) 인천광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
중구	1425	1.2	33436	0.1	34861	0.1	23773	0.1	4.1
동구	1552	0.9	35875	0.2	37427	0.2	25795	0.3	4.1
남구	8828	6.0	195194	0.8	204022	0.9	144029	1.2	4.3
연수구	4796	2.1	113270	0.4	118066	0.4	80534	0.5	4.1
남동구	7740	4.4	178231	0.5	185971	0.5	128071	0.6	4.2
부평구	10408	3.6	241996	0.8	252404	0.8	173949	1.1	4.1
계양구	5985	3.9	142544	0.5	148529	0.5	100083	0.7	4.0
서구	6388	2.6	151048	0.3	157436	0.3	105509	0.4	4.1
강화군	1140	0.7	30864	0.1	32004	0.1	20741	0.2	3.6
계	48,262		1,122,458		1,170,720		802,484		4.12

(4) 대구광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
중구	1095	5.1	39397	0.3	40492	0.3	32301	0.4	2.7
동구	4110	3.9	148932	0.8	153042	0.8	119247	1.0	2.7
서구	3568	8.8	122192	0.5	125760	0.4	100035	0.5	2.8
남구	2448	3.0	85661	0.4	88109	0.5	71291	0.6	2.8
북구	4670	3.1	173277	0.7	177947	0.7	138379	0.9	2.6
수성구	5347	4.6	193877	0.6	199224	0.6	158302	0.7	2.7
달서구	7082	6.1	253162	0.6	260244	0.6	202253	0.7	2.7
달성군	1664	3.5	64004	0.5	65668	0.5	48617	0.7	2.5
계	29,984		1,080,502		1,110,486		870,425		2.70

(5) 광주광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
동구	1145	-	48052	0.8	49197	0.8	43477	0.9	2.3
서구	2963	9.7	116842	1.0	119805	1.0	101861	1.1	2.5
남구	2398	9.5	92121	0.8	94519	0.8	83841	0.9	2.5
북구	4771	8.1	189445	0.9	194216	0.9	168929	1.0	2.5
광산구	2244	4.3	99006	0.9	101250	0.9	81435	1.1	2.2
계	13,521		545,466		558,987		479,543		2.42

(6) 대전광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
동구	3517	7.5	106698	0.6	110215	0.6	91490	0.7	3.2
중구	3668	6.9	113872	0.7	117540	0.8	95300	0.9	3.1
서구	6503	6.7	202089	1.0	208592	1.0	171177	1.2	3.1
유성구	2401	4.8	74904	1.0	77305	1.0	61542	1.3	3.1
대덕구	3126	5.3	98123	0.6	101249	0.7	80456	0.8	3.1
계	19,215		595,686		614,901		499,965		3.12

(7) 울산광역시

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
중구	2999	12.5	102182	0.9	105181	0.8	75651	1.1	2.9
남구	4121	7.9	145864	0.8	149985	0.9	106917	1.3	2.7
동구	2307	5.8	79800	0.7	82106	0.7	57752	1.0	2.8
울주군	1964	12.5	73038	0.7	75003	0.8	52060	1.1	2.6
계	11,391		400,884		412,275		292,380		2.76

(8) 경기도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
평택시	2517	19.4	127633	1.2	130150	1.0	70197	1.6	1.9
용인시	3116	11.0	165547	1.4	168663	1.3	92847	2.2	1.8
파주시	1521	20.6	60879	0.7	62400	0.5	40725	0.8	2.4
이천시	1280	12.3	66653	1.1	67932	1.0	38897	1.6	1.9
안성시	1079	3.1	53890	0.4	54969	0.4	30370	0.7	2.0
광주시	1474	19.6	47258	1.0	48732	1.1	28978	1.8	3.0
포천군	1951	18.3	69137	1.7	71087	1.8	39858	3.1	2.7
가평/양평군 가평군 양평군	1735 701 1034	21.7	66261 26769 39492	2.5	67997 27471 40526	2.6	38632 15607 23025	4.2	2.6
수원시	12930	8.8	448642	0.8	461572	0.9	366422	1.1	2.8
성남시	12850	4.0	443162	1.0	456012	1.0	362067	1.2	2.8
안양시	8196	7.7	284369	0.5	292565	0.6	231573	0.7	2.8
부천시	10331	5.0	364781	0.8	375112	0.8	296797	1.0	2.8
동두천시	956	1.3	35180	0.1	36137	0.1	27651	0.1	2.6
고양시	9551	5.1	363163	0.8	372714	0.8	282394	1.0	2.6
과천/구리 과천시 구리시	3182 885 2297	2.3	116686 32439 84247	0.4	119868 33323 86545	0.4	92595 25741 66854	0.5	2.7
군포시	3430	2.6	122553	0.4	125982	0.4	98143	0.5	2.7
하남시	1692	1.0	59620	0.4	61312	0.4	47443	0.4	2.8
김포시	1007	50.0	74935	4.2	75942	4.1	55106	6.0	1.3
화성군	1549	17.8	62792	0.7	64341	0.6	31625	1.1	2.4
여주/연천군 여주군 연천군	1307 874 433	16.1	67857 45396 22461	0.6	69164 46271 22893	0.8	39157 26196 12961	1.4	1.9
의정부시	5394	4.7	167699	0.8	173093	0.7	130821	0.9	3.1
광명시	5205	4.4	154093	0.7	159298	0.7	122313	0.9	3.3
안산시	9986	12.4	254919	1.3	264905	1.4	193662	1.8	3.8
남양주시	5225	5.3	166283	0.6	171508	0.7	125438	0.9	3.0
오산시	1761	2.3	50428	0.4	52188	0.4	39387	0.5	3.4
시흥시	4490	5.8	142039	1.7	146529	1.8	106261	2.4	3.1
의왕시	1833	1.2	56088	0.3	57922	0.3	43242	0.4	3.2
양주군	2150	16.5	37886	1.9	40035	1.0	25862	1.6	5.4
계	117,698		4,130,433		4,248,129		3,097,463		2.77

(9) 강원도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
춘천시	3014	7.5	107914	1.4	110928	1.4	94725	1.6	2.7
원주시	3042	9.7	115911	1.2	118952	1.3	98483	1.5	2.6
강릉시	2587	9.4	99660	1.2	102248	1.2	84801	1.3	2.5
동해시	1060	7.1	41899	1.0	42959	0.9	34867	1.0	2.5
태백시	531	-	24334	0.9	24865	0.8	19769	1.0	2.1
속초시	923	3.0	37886	0.7	38810	0.7	31589	0.8	2.4
삼척시	805	2.7	33616	1.3	34421	1.2	27063	1.5	2.3
홍천군	213	-	31845	1.6	32058	1.5	16399	2.7	0.7
횡성/영월 /고성군	443	24.4	55835	1.3	56278	1.3	28900	2.3	0.8
횡성군	153		19263		19416		9971		
영월군	174		21943		22117		11358		
고성군	116		14629		14745		7571		
평창/정선 /양양군	390	-	55837	2.0	56227	2.0	28426	3.6	0.7
평창군	144		20660		20804		10518		
정선군	152		21776		21929		11086		
양양군	94		13401		13494		6822		
철원/화천군	220	-	32320	0.7	32541	0.7	16811	1.3	0.7
철원군	149		21848		21998		11364		
화천군	71		10472		10543		5447		
양구/인제군	167	20.3	24044	0.6	24210	0.6	12129	1.1	0.7
양구군	69		9930		9999		5009		
인제군	98		14114		14211		7120		
계	13,395		661,101		674,497		493,962		1.99

(10) 충청북도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
청주시	4502	13.5	249030	1.7	253532	1.7	232033	1.8	1.8
충주시	1505	8.8	95231	1.5	96735	1.4	81200	1.6	1.6
제천시	1015	9.9	62193	1.2	63208	1.2	53008	1.4	1.6
청원군	675	10.0	51811	1.3	52485	1.3	30957	2.1	1.3
보은/영동군	534	10.7	47612	0.8	48145	0.9	25532	1.6	1.1
보은군	226		20187		20413		10826		
영동군	308		27425		27732		14706		
옥천/진천군	627	16.8	53367	1.5	53994	1.5	30558	2.5	1.2
옥천군	309		26310		26619		15065		
진천군	318		27057		27375		15493		
괴산군	313	-	31662	1.0	31974	1.0	17192	1.7	1.0
음성군	437	9.1	38892	1.1	39329	1.0	20909	1.8	1.1
단양군	179	5.1	17075	0.6	17255	0.5	9090	0.9	1.0
계	9,787		646,873		656,657		500,479		1.49

(11) 충청남도

시군구	실업자 수		취업자 수		경찰인구		비경찰인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
천안시	4968	6.8	195570	1.2	200538	1.2	115107	1.9	2.5
공주시	1470	6.5	61740	0.9	63210	0.7	36615	1.2	2.3
보령시	1161	3.3	51960	0.5	53121	0.5	29339	0.8	2.2
아산시	2130	6.4	84895	0.7	87025	0.8	51174	1.3	2.4
서산시	1552	8.1	67695	0.8	69247	0.8	37978	1.3	2.2
논산시	1740	-	77658	0.9	79398	0.9	44719	1.5	2.2
금산/서천/청양군	931	12.1	99669	1.1	100601	1.1	41570	2.2	0.9
금산군	334		35781		36116		14924		
청양군	211		22625		22836		9436		
서천군	386		41263		41649		17210		
연기/홍성군	1208	35.5	97747	1.5	98955	1.7	48169	3.0	1.2
연기군	563		45550		46113		22447		
홍성군	645		52197		52842		25722		
부여군	463	8.3	51133	1.1	51596	1.1	21207	2.3	0.9
예산/태안군	1049	14.2	97785	2.0	98834	1.8	43141	3.5	1.1
예산군	626		58378		59004		25755		
태안군	423		39407		39830		17386		
당진군	733	6.6	68579	1.7	69311	1.7	30366	3.4	1.1
계	17,405		954,431		971,836		499,385		1.79

(12) 전라북도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
전주시	6842	9.9	248338	0.8	255180	0.8	220126	0.9	2.7
군산시	2959	7.7	113537	1.1	116496	1.2	95332	1.4	2.5
익산시	3317	10.0	131769	1.6	135086	1.5	110495	1.8	2.5
정읍시	1311	4.5	56217	0.9	57528	0.9	44651	1.1	2.3
남원시	993	6.0	41760	0.9	42754	0.7	33064	0.9	2.3
김제시	1082	3.1	46381	0.9	47463	0.9	36658	1.1	2.3
완주군	454	-	39679	0.5	40133	0.5	26227	0.6	1.1
진안/장수 /임실군	497	17.1	45803	2.4	46300	2.3	26595	3.6	1.1
진안군	162		14886		15048		8643		
장수군	151		13878		14029		8058		
임실군	184		17039		17223		9894		
무주/순창 /고창군	533	-	64664	1.6	65197	1.6	36623	2.6	0.8
무주군	117		14161		14278		8020		
순창군	131		15843		15973		8973		
고창군	285		34660		34946		19630		
부안군	433	25.5	33699	1.1	34132	1.9	19759	2.9	1.3
계	18,421		821,847		840,269		649,530		2.19

(13) 전라남도

시군구	실업자 수		취업자 수		경찰인구		비경찰인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
목포시	4541	8.3	118683	1.3	123224	1.2	94844	1.5	3.7
여수시	4592	8.0	149653	1.1	154245	1.0	111162	1.4	3.0
순천시	3929	5.5	129593	1.5	133523	1.4	97613	1.9	2.9
나주시	1442	2.2	50927	0.9	52369	0.9	37064	1.3	2.8
광양시	1850	3.4	62001	1.0	63851	1.0	44429	1.4	2.9
담양군	339	7.0	29632	0.2	29972	0.3	11993	0.7	1.1
곡성/구례 /보성군	705	7.0	68297	0.3	69001	0.3	25065	0.7	1.0
곡성군	209		20216		20424		7419		
구례군	177		17143		17319		6291		
보성군	319		30938		31258		11355		
고흥군	511	15.8	50019	0.7	50529	0.6	17548	1.4	1.0
화순/장흥군	758	35.0	65054	1.0	65812	0.9	24655	2.0	1.2
화순군	459		39358		39816		14916		
장흥군	299		25696		25996		9739		
강진군	261	7.1	25434	0.1	25695	0.1	9341	1.1	1.0
해남군	512	6.7	49068	1.0	49580	0.9	18013	2.3	1.0
영암군	341	-	32595	0.5	32936	0.5	12273	1.5	1.0
무안군	367	-	30689	0.2	31056	0.2	12587	0.3	1.2
함평군	234	-	23136	0.5	23370	0.5	8558	1.0	1.0
영광/완도 /진도군	919	-	89006	1.3	89924	1.3	31925	3.3	1.0
영광군	365		35335		35700		12674		
완도군	340		32932		33272		11812		
진도군	214		20739		20952		7439		
장성군	289	-	27245	0.7	27534	0.7	10406	1.6	1.0
계	21,590		1,001,032		1,022,621		567,476		2.11

(14) 경상북도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
포항시	5929	5.8	245605	0.9	251534	0.8	151399	1.3	2.4
경주시	3095	5.2	134990	0.5	138085	0.5	82527	0.8	2.2
김천시	1746	6.3	74884	0.4	76630	0.3	46328	0.4	2.3
안동시	2124	2.1	91545	0.7	93669	0.6	56802	1.0	2.3
구미시	4240	8.0	156086	0.7	160326	0.7	101777	1.1	2.6
영주시	1443	2.5	63242	0.2	64685	0.3	38537	0.4	2.2
영천시	1219	2.0	57258	0.5	58478	0.5	33324	0.8	2.1
상주시	1243	1.8	60660	0.5	61903	0.5	35015	0.8	2.0
문경시	1074	4.1	46918	0.3	47992	0.4	28666	0.6	2.2
경산시	2822	5.1	107435	0.3	110257	0.3	71137	0.4	2.6
군위/영양 /성주군	191	-	58327	1.5	58518	1.5	25986	2.9	0.3
군위군	61		18606		18667		8290		
영양군	40		12074		12113		5379		
성주군	90		27647		27738		12317		
의성/청송군	182	-	60233	0.7	60415	0.7	25230	1.5	0.3
의성군	125		41501		41626		17383		
청송군	57		18732		18789		7847		
영덕/청도군	263	49.2	54295	2.1	54558	2.3	24912	4.4	0.5
영덕군	129		26659		26788		12232		
청도군	134		27636		27770		12680		
고령/봉화군	141	-	42784	0.6	42925	0.6	19087	1.2	0.3
고령군	67		20322		20389		9066		
봉화군	74		22462		22536		10021		
칠곡/예천 /울진군	605	29.5	113434	2.1	114039	2.1	53701	3.8	0.5
칠곡군	280		52520		52800		24864		
예천군	152		28585		28738		13533		
울진군	173		32329		32501		15304		
계	26,317		1,367,696		1,394,014		794,428		1.89

(15) 경상남도

시군구	실업자 수		취업자 수		경찰인구		비경찰인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
창원시	7542	4.8	232808	0.8	240350	0.7	170936	1.0	3.1
마산시	7032	6.2	211351	0.8	218382	0.7	160459	1.0	3.2
진주시	5128	2.5	160652	0.8	165780	0.7	122189	1.0	3.1
진해시	1987	2.3	61677	0.8	63663	0.7	44409	1.0	3.1
통영시	1939	2.2	60948	0.3	62887	0.2	43144	0.4	3.1
사천시	1729	1.4	55311	0.6	57040	0.6	39019	0.9	3.0
김해시	5017	5.4	156036	0.9	161053	0.7	111178	1.1	3.1
밀양시	1820	8.3	58637	0.3	60457	0.3	41017	0.4	3.0
거제시	2587	3.0	80124	0.5	82711	0.5	56463	0.7	3.1
양산시	2883	4.8	91041	0.4	93924	0.4	64664	0.7	3.1
의령/합천군	292	-	31922	1.9	32214	1.9	16962	2.3	0.9
의령군	105		11492		11597		6106		
합천군	187		20430		20617		10856		
함안/하동/거창군	862	21.1	68388	2.0	69251	2.1	34933	4.1	1.2
함안군	292		23184		23476		11842		
하동군	266		21132		21399		10794		
거창군	304		24072		24376		12297		
창녕/고성군	445	-	47718	2.8	48163	2.8	23594	5.6	0.9
창녕군	237		25434		25671		12576		
고성군	208		22284		22492		11018		
산청/함양/남해군	562	22.4	53068	0.7	53631	0.8	25756	1.5	1.0
산청군	157		14806		14963		7186		
함양군	178		16769		16947		8139		
남해군	227		21493		21721		10431		
계	39,825		1,369,681		1,409,506		954,723		2.83

(16) 제주도

시군구	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
제주시	1719	20.3	149694	0.8	151413	0.7	80566	1.3	1.1
서귀포시	455	10.4	45293	1.5	45747	1.5	22778	3.0	1.0
북제주군	0	-	46681	1.8	46681	1.8	15555	4.9	-
남제주군	0	-	37599	1.4	37599	1.4	12026	3.9	-
계									

부록2 시·도단위의 세부영역 추정결과(전국)

2002년 10월의 경제활동인구조사 자료로부터 대영역의 “성별(남, 여)-연령대별(15-29세, 30-39세, 40-49세, 50세 이상)” 범주와 “성별(남, 여)-교육정도별(초등졸 이하, 중졸, 고졸, 대졸 이상)” 범주에 대한 “실업자 수”, “취업자 수”, “경제활동인구”, “비경제활동인구”의 복합추정법에 의한 추정 결과를 요약하면 다음과 같다.

(1) “성별-연령대별” 범주의 추정결과

① 서울특별시

구 분	실업자 수		취업자 수		경제활동인구		비경제활동인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	41550	5.1	521752	1.4	563302	1.3	659232	1.0	7.38
	30-39세	25555	7.4	864309	0.4	889865	0.3	43192	5.6	2.87
	40-49세	14948	7.5	753294	0.3	768243	0.2	34401	4.8	1.95
	50세 이상	7460	11.6	603820	1.0	611280	1.0	318162	1.7	1.22
여	15-29세	25303	6.2	508308	1.4	533611	1.3	684855	0.9	4.74
	30-39세	8065	11.3	438635	1.2	446700	1.2	406185	1.2	1.81
	40-49세	6116	10.6	442031	1.3	448146	1.3	354181	1.5	1.36
	50세 이상	3428	20.8	315524	1.8	318952	1.8	684111	0.8	1.07
계	132,425		4,447,673		4,580,099		3,184,319		2.89	

② 부산광역시

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	14910	3.7	168507	0.9	183417	0.8	280761	0.5	8.13
	30-39세	8776	4.1	279145	0.2	287922	0.2	16072	2.9	3.05
	40-49세	6629	5.6	290904	0.3	297533	0.2	15854	3.9	2.23
	50세이상	3662	6.7	239916	0.6	243578	0.6	144342	1.0	1.50
여	15-29세	8343	3.2	154578	0.8	162921	0.7	272436	0.4	5.12
	30-39세	2712	6.2	149751	0.8	152463	0.7	149101	0.7	1.78
	40-49세	2936	6.0	179753	0.7	182689	0.7	143999	0.9	1.61
	50세이상	1234	15.1	128910	1.2	130144	1.1	317021	0.5	0.95
계		49,202		1,591,464		1,640,667		1,339,586		3.00

③ 인천광역시

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	13831	4.0	113964	0.7	127795	0.6	173859	0.4	10.82
	30-39세	10447	3.6	239196	0.2	249644	0.1	12446	1.9	4.18
	40-49세	6882	4.7	220177	0.2	227059	0.1	10621	2.1	3.03
	50세이상	2908	7.5	132019	0.5	134926	0.5	74249	0.7	2.16
여	15-29세	7613	2.8	102611	0.7	110224	0.6	172805	0.3	6.91
	30-39세	2963	4.3	123776	0.6	126739	0.6	112215	0.6	2.34
	40-49세	2752	5.4	120508	0.5	123260	0.4	87045	0.6	2.23
	50세이상	867	8.5	70207	0.8	71074	0.8	159242	0.3	1.22
계		48,263		1,122,458		1,170,721		802,482		4.12

④ 대구광역시

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	9120	2.7	117468	0.7	126587	0.7	187561	0.4	7.20
	30-39세	5712	2.5	209405	0.2	215117	0.1	11752	2.6	2.66
	40-49세	3953	4.0	194948	0.2	198902	0.1	10608	2.4	1.99
	50세이상	1941	5.1	142463	0.4	144404	0.4	87407	0.6	1.34
여	15-29세	5225	3.1	109381	0.6	114606	0.6	180944	0.3	4.56
	30-39세	1777	6.8	112545	0.6	114322	0.6	108455	0.6	1.55
	40-49세	1634	6.7	115260	0.6	116893	0.6	90336	0.7	1.40
	50세이상	623	-	79032	1.1	79655	1.1	193362	0.4	0.78
계		29,985		1,080,502		1,110,486		870,425		2.70

⑤ 광주광역시

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	4416	1.7	66535	0.4	70951	0.4	109209	0.3	6.22
	30-39세	2520	1.5	109547	0.1	112067	0.1	6354	1.4	2.25
	40-49세	1519	1.7	91638	0.1	93157	0.1	5028	1.5	1.63
	50세이상	746	3.5	67043	0.2	67789	0.2	43290	0.4	1.10
여	15-29세	2612	1.7	64097	0.4	66709	0.4	111446	0.2	3.92
	30-39세	790	1.1	56973	0.4	57764	0.4	61310	0.4	1.37
	40-49세	661	2.9	53289	0.3	53950	0.3	46825	0.4	1.23
	50세이상	256	5.6	36343	0.7	36599	0.7	96080	0.3	0.70
계		13,520		545,465		558,986		479,542		2.42

⑥ 대전광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	6194	1.6	70788	0.5	76982	0.4	112856	0.3	8.05
	30-39세	3596	1.7	119665	0.1	123261	0.1	6710	1.6	2.92
	40-49세	2389	2.2	108002	0.1	110391	0.1	6039	1.6	2.16
	50세이상	1081	1.7	71876	0.2	72956	0.2	46556	0.3	1.48
여	15-29세	3482	2.1	64608	0.4	68090	0.4	112761	0.3	5.11
	30-39세	1150	2.9	62241	0.4	63392	0.4	64671	0.4	1.81
	40-49세	976	5.9	60717	0.3	61693	0.3	52138	0.3	1.58
	50세이상	347	-	37790	0.4	38137	0.4	98234	0.2	0.91
계	19,215		595,687		614,902		499,965		3.12	

⑦ 울산광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	3478	1.4	48407	0.4	51885	0.3	76645	0.2	6.70
	30-39세	2750	0.8	106737	0.1	109487	0.1	5569	1.7	2.51
	40-49세	1782	2.2	96445	0.1	98227	0.1	4641	1.4	1.81
	50세이상	681	2.1	51224	0.2	51905	0.2	29776	0.4	1.31
여	15-29세	1891	1.5	42818	0.3	44709	0.3	73101	0.2	4.23
	30-39세	787	5.1	54645	0.2	55432	0.3	50738	0.2	1.42
	40-49세	715	5.2	52725	0.3	53440	0.3	37785	0.3	1.34
	50세이상	167	-	25191	0.7	25358	0.7	59135	0.3	0.66
계	12,251		478,192		490,443		337,390		2.50	

⑧ 경기도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	40739	6.3	428065	1.7	468804	1.6	586059	1.4	8.69
	30-39세	22114	8.2	887185	0.3	909298	0.2	42297	5.8	2.43
	40-49세	12877	11.0	749812	0.4	762689	0.4	50934	5.9	1.69
	50세이상	8082	15.1	515117	1.0	523199	1.0	265195	2.2	1.54
여	15-29세	18547	11.6	349611	2.0	368159	1.9	715590	1.0	5.04
	30-39세	6628	12.5	440223	1.6	446851	1.6	524274	1.4	1.48
	40-49세	6778	17.0	445797	1.2	452574	1.2	311857	2.0	1.50
	50세이상	1933	22.2	324278	1.8	326211	1.8	610968	1.0	0.59
계	117,698		4,140,088		4,257,785		3,107,174		2.76	

⑨ 강원도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	5322	6.2	51582	2.1	56903	2.0	104143	1.1	9.35
	30-39세	2697	6.5	107897	0.4	110594	0.3	11625	3.2	2.44
	40-49세	1623	8.6	110372	0.4	111995	0.4	9739	4.3	1.45
	50세이상	972	8.2	119572	0.8	120544	0.7	54213	1.6	0.81
여	15-29세	1534	9.7	39842	2.1	41376	2.1	104351	0.8	3.71
	30-39세	343	20.3	60675	1.2	61019	1.2	52756	1.4	0.56
	40-49세	699	4.6	71206	1.0	71905	1.0	43774	1.7	0.97
	50세이상	205	22.2	99955	1.3	100160	1.3	113360	1.1	0.20
계	13,395		661,101		674,496		493,961		1.99	

⑩ 충청북도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	3944	4.0	55780	2.2	59724	2.1	112190	1.2	6.60
	30-39세	1963	4.4	109358	0.5	111321	0.5	11722	4.9	1.76
	40-49세	1102	6.2	106225	0.4	107327	0.3	9241	4.1	1.03
	50세이상	664	13.9	109284	0.7	109948	0.6	50285	1.4	0.60
여	15-29세	1240	9.9	45585	3.2	46824	3.1	113844	1.3	2.65
	30-39세	257	-	61019	1.4	61277	1.4	54966	1.6	0.42
	40-49세	469	14.1	67195	1.0	67664	1.0	41968	1.6	0.69
	50세이상	148	38.5	92426	1.1	92574	1.1	106265	0.9	0.16
계		9,787		646,872		656,659		500,481		1.49

⑪ 충청남도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	6450	2.3	98771	0.7	105221	0.7	94553	0.5	6.13
	30-39세	3009	4.5	159050	0.2	162059	0.2	6042	2.9	1.86
	40-49세	1865	3.9	145129	0.2	146994	0.2	7701	2.6	1.27
	50세이상	1497	5.8	176313	0.5	177810	0.5	62185	1.1	0.84
여	15-29세	2470	3.5	62395	1.1	64865	1.0	101538	0.5	3.81
	30-39세	821	5.1	77725	0.5	78546	0.5	53620	0.6	1.05
	40-49세	902	4.8	96630	0.4	97532	0.4	35826	0.8	0.92
	50세이상	390	6.1	138419	1.1	138809	1.1	137921	0.8	0.28
계		17,404		954,432		971,836		499,386		1.79

⑫ 전라북도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	7539	8.1	66975	2.1	74514	1.9	141978	1.0	10.12
	30-39세	3589	9.4	128344	0.6	131934	0.5	13541	5.1	2.72
	40-49세	1957	11.6	123042	0.5	125000	0.4	10548	5.2	1.57
	50세이상	1403	14.1	155903	0.9	157307	0.9	72165	2.0	0.89
여	15-29세	2211	10.9	55288	2.3	57499	2.2	135709	1.0	3.85
	30-39세	499	25.8	72470	1.7	72969	1.6	64638	1.9	0.68
	40-49세	917	18.5	82554	1.3	83471	1.3	52245	2.0	1.10
	50세이상	305	-	137271	1.5	137576	1.5	158706	1.3	0.22
계		18,420		821,847		840,270		649,530		2.19

⑬ 전라남도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	8450	9.1	70883	2.5	79333	2.1	112738	1.1	10.65
	30-39세	4467	10.7	147863	0.6	152330	0.5	12176	5.2	2.93
	40-49세	2618	15.0	143584	0.5	146201	0.4	9806	4.8	1.79
	50세이상	1759	12.1	211154	0.9	212914	0.9	72152	2.0	0.83
여	15-29세	2259	12.4	49995	2.0	52254	2.1	107412	0.8	4.32
	30-39세	503	30.7	81028	1.5	81532	1.5	52680	1.8	0.62
	40-49세	1115	22.2	94904	1.3	96019	1.2	42870	2.1	1.16
	50세이상	419	-	201620	1.6	202038	1.6	157641	1.5	0.21
계		21,590		1,001,031		1,022,621		567,475		2.11

⑭ 경상북도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	9708	3.6	139996	1.1	149705	1.1	149048	0.9	6.48
	30-39세	4368	4.3	227612	0.2	231981	0.2	9159	3.3	1.88
	40-49세	2851	5.7	209209	0.2	212061	0.2	11766	3.0	1.34
	50세이상	2244	7.8	243645	0.7	245888	0.6	93987	1.4	0.91
여	15-29세	3831	4.8	93940	1.2	97771	1.1	162591	0.6	3.92
	30-39세	1300	8.2	113108	0.7	114408	0.6	89236	0.7	1.14
	40-49세	1434	7.7	142134	0.5	143568	0.5	60517	1.1	1.00
	50세이상	583	13.3	198049	1.2	198632	1.2	218123	0.9	0.29
계		26,319		1,367,693		1,394,014		794,427		1.89

⑮ 경상남도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	15-29세	14173	4.4	141232	1.1	155405	1.0	186135	0.9	9.12
	30-39세	6963	5.9	251759	0.3	258722	0.2	12114	4.8	2.69
	40-49세	4565	11.4	228615	0.4	233180	0.3	15079	4.3	1.96
	50세이상	2984	7.3	207171	0.7	210156	0.7	96387	1.6	1.96
여	15-29세	5673	4.4	98196	1.4	103869	1.4	204327	0.7	5.46
	30-39세	2266	13.9	127835	1.0	130102	1.0	126407	1.0	1.74
	40-49세	2379	12.7	149008	0.7	151386	0.7	82236	1.3	1.57
	50세이상	821	14.4	165867	1.1	166687	1.1	232039	0.8	0.49
계		39,824		1,369,683		1,409,507		954,724		2.83

(16) 제주도

구분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	15-29세	811	0.6	30996	0.2	31807	0.2	26082	0.2	2.55
	30-39세	380	-	53480	0.1	53860	0.1	1702	1.5	0.71
	40-49세	217	1.4	43429	0.1	43646	0.1	1868	0.7	0.50
	50세이상	164	-	41275	0.1	41439	0.1	12244	0.3	0.40
여	15-29세	329	-	21947	0.3	22276	0.3	29531	0.1	1.48
	30-39세	118	2.0	26529	0.2	26647	0.2	17740	0.2	0.44
	40-49세	113	1.7	28351	0.2	28464	0.1	10553	0.3	0.40
	50세이상	43	-	33259	0.3	33302	0.3	31205	0.2	0.13
계	2,175		279,266		281,441		130,925		0.77	

(2) “성별-교육정도별” 범주의 추정결과

① 서울특별시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	5151	15.6	153529	1.7	158680	1.7	140131	1.7	3.25
	중졸	7952	12.0	271043	1.1	278995	1.1	270749	1.0	2.85
	고졸	49931	5.0	1393625	0.6	1443556	0.5	531301	1.3	3.46
	대졸이상	26210	6.3	919798	0.5	946008	0.5	119257	3.7	2.77
여	초졸이하	1129	23.6	203369	1.9	204498	1.9	456160	0.8	0.55
	중졸	3968	10.8	244012	1.5	247981	1.5	390875	0.9	1.60
	고졸	26085	7.4	879486	1.0	905571	1.0	989265	0.8	2.88
	대졸이상	11999	9.2	382810	1.3	394809	1.3	286580	1.6	3.04
계	132,425		4,447,672		4,580,098		3,184,318		2.89	

② 부산광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	2673	8.7	78774	1.0	81447	1.0	74868	1.1	3.28
	중졸	4138	6.5	138896	0.6	143034	0.6	127063	0.6	2.89
	고졸	21461	3.1	560822	0.3	582283	0.3	228016	0.7	3.69
	대졸이상	5752	2.6	200204	0.2	205956	0.2	27111	1.5	2.79
여	초졸이하	646	14.7	106116	1.2	106761	1.2	251539	0.5	0.61
	중졸	2296	11.0	116588	1.0	118883	1.0	186348	0.6	1.93
	고졸	9481	3.0	311846	0.6	321326	0.5	383889	0.4	2.95
	대졸이상	2756	4.2	78219	0.5	80976	0.5	60751	0.6	3.40
계	49,203		1,591,465		1,640,666		1,339,585		3.00	

③ 인천광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	2579	7.4	54808	0.7	57406	0.7	46538	0.7	4.49
	중졸	3823	6.2	85729	0.5	89552	0.5	74942	0.5	4.27
	고졸	22491	3.0	436581	0.2	459072	0.2	136884	0.5	4.90
	대졸이상	5161	2.4	128136	0.1	133297	0.1	13054	0.9	3.87
여	초졸이하	571	19.0	66211	0.9	66782	0.9	138653	0.4	0.86
	중졸	1996	4.7	71525	0.5	73522	0.5	105073	0.3	2.71
	고졸	9728	2.6	239660	0.4	249388	0.4	259084	0.4	3.90
	대졸이상	1895	3.0	39807	0.3	41702	0.3	28255	0.4	4.54
계	48,244		1,122,457		1,170,721		802,483		4.12	

④ 대구광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1718	7.6	58307	0.8	60025	0.8	52513	0.9	2.86
	중졸	2410	6.6	85380	0.4	87789	0.4	79677	0.4	2.75
	고졸	12671	2.1	368568	0.3	381239	0.3	146170	0.6	3.32
	대졸이상	3908	2.1	152178	0.2	156086	0.1	19046	1.1	2.50
여	초졸이하	402	-	74625	1.2	75027	1.1	165243	0.5	0.54
	중졸	1370	12.8	74598	0.8	75969	0.8	114416	0.5	1.80
	고졸	5609	3.0	206399	0.5	212007	0.5	247850	0.4	2.65
	대졸이상	1898	2.2	60446	0.4	62343	0.4	45511	0.5	3.04
계	29,986		1,080,501		1,110,485		870,426		2.70	

⑤ 광주광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	601	3.7	22988	0.4	23589	0.4	24158	0.4	2.55
	중졸	832	2.8	32033	0.3	32865	0.2	43625	0.2	2.53
	고졸	5516	1.2	181048	0.1	186565	0.1	82654	0.3	2.96
	대졸이상	2254	1.7	98798	0.1	101052	0.1	13474	0.8	2.23
여	초졸이하	145	11.1	31138	0.7	31283	0.7	79069	0.3	0.46
	중졸	484	2.9	30873	0.4	31358	0.4	61044	0.2	1.54
	고졸	2678	1.5	112457	0.3	115135	0.3	147157	0.2	2.32
	대졸이상	1011	1.6	36130	0.3	37141	0.3	28362	0.4	2.72
계	13,521		545,465		558,988		479,543		2.42	

⑥ 대전광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1032	3.5	30427	0.4	31459	0.4	29113	0.4	3.28
	중졸	1272	2.9	37728	0.3	39000	0.3	44754	0.3	3.26
	고졸	7588	1.5	189400	0.2	196988	0.2	84503	0.4	3.85
	대졸이상	3371	1.6	112890	0.1	116262	0.1	13835	0.9	2.90
여	초졸이하	221	-	35855	0.5	36077	0.5	84926	0.2	0.61
	중졸	686	5.9	32964	0.4	33650	0.4	61643	0.2	2.04
	고졸	3697	2.6	117514	0.3	121211	0.3	149413	0.2	3.05
	대졸이상	1348	2.3	38906	0.3	40254	0.3	31778	0.4	3.35
계	19,215		595,684		614,901		499,965		3.12	

⑦ 울산광역시

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	618	-	21593	0.5	22212	0.4	19850	0.4	2.78
	중졸	947	5.3	36050	0.4	36998	0.3	33453	0.3	2.56
	고졸	5853	0.9	193736	0.1	199589	0.1	58599	0.3	2.93
	대졸이상	1276	0.6	51407	0.1	52683	0.1	4759	0.3	2.42
여	초졸이하	142	12.8	27076	0.7	27218	0.7	56283	0.3	0.52
	중졸	531	5.1	31653	0.4	32185	0.3	45340	0.2	1.65
	고졸	2355	1.9	98686	0.2	101041	0.2	106694	0.2	2.33
	대졸이상	529	1.3	17990	0.2	18518	0.2	12413	0.2	2.86
계	12,251		478,191		490,444		337,391		2.50	

⑧ 경기도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	2752	10.6	233128	1.2	235881	1.2	168162	1.8	1.17
	중졸	6955	14.9	282134	1.1	289089	1.0	276901	1.2	2.41
	고졸	56632	5.6	1429874	0.6	1486507	0.5	431291	2.0	3.81
	대졸이상	17882	9.4	639015	0.6	656897	0.5	83755	4.6	2.72
여	초졸이하	1453	27.2	314072	1.4	315524	1.4	485609	1.0	0.46
	중졸	4859	13.8	235773	1.8	240632	1.8	390002	1.2	2.02
	고졸	21567	10.5	778939	1.3	800506	1.3	1055180	1.0	2.69
	대졸이상	5595	18.3	227153	1.9	232748	1.9	216271	2.3	2.40
계	117,695		4,140,088		4,257,784		3,107,171		2.76	

⑨ 강원도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	초졸이하	777	13.6	82457	0.8	83234	0.8	42003	1.6	0.93
	중졸	1206	9.7	56916	1.0	58122	1.0	46711	1.2	2.07
	고졸	6931	5.5	183817	0.7	190748	0.7	83037	1.5	3.63
	대졸이상	1738	7.0	66995	0.5	68732	0.5	9276	3.5	2.53
여	초졸이하	634	7.2	114078	1.1	114711	1.1	110225	1.1	0.55
	중졸	75	-	41205	1.3	41279	1.3	61868	0.9	0.18
	고졸	1503	10.1	93146	1.3	94648	1.3	127260	1.0	1.59
	대졸이상	532	12.5	22489	1.4	23021	1.4	13584	2.3	2.31
계		13,396		661,103		674,495		493,964		1.99

⑩ 충청북도

구 분		실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)
		추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	
남	초졸이하	546	12.0	79852	0.9	80398	0.9	41248	1.7	0.68
	중졸	805	6.8	49334	0.9	50139	0.8	45630	0.9	1.61
	고졸	5102	3.8	184816	0.8	189919	0.8	87864	1.7	2.69
	대졸이상	1256	5.3	67461	0.4	68717	0.4	9444	2.8	1.83
여	초졸이하	449	18.7	106703	1.0	107152	1.0	104698	1.0	0.42
	중졸	56	-	38810	1.5	38866	1.5	61153	1.0	0.14
	고졸	1178	9.8	97718	1.8	98896	1.8	137111	1.3	1.19
	대졸이상	395	10.2	22178	1.5	22573	1.4	13332	2.5	1.75
계		9,787		646,872		656,660		500,480		1.49

⑪ 충청남도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1372	10.7	143163	0.7	144535	0.7	54700	1.3	0.95
	중졸	1260	6.0	78684	0.4	79944	0.4	45287	0.5	1.58
	고졸	8654	2.0	280654	0.2	289309	0.2	63805	0.8	2.99
	대졸이상	1562	2.1	76539	0.1	78101	0.1	6385	1.2	2.00
여	초졸이하	446	6.7	167412	0.9	167858	0.9	140630	0.8	0.27
	중졸	860	5.1	55425	0.7	56285	0.7	58899	0.5	1.53
	고졸	2870	3.3	131398	0.5	134267	0.5	117203	0.4	2.14
	대졸이상	381	5.4	21156	0.5	21537	0.4	12477	0.5	1.77
계	17,405		954,431		971,836		499,386		1.79	

⑫ 전라북도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	986	9.2	101824	1.0	102810	0.9	53310	1.9	0.96
	중졸	1545	11.7	62068	1.0	63614	1.0	61288	1.1	2.43
	고졸	9394	6.9	220618	0.8	230012	0.7	110504	1.5	4.08
	대졸이상	2647	11.8	91627	0.8	94273	0.7	14191	4.6	2.81
여	초졸이하	699	18.0	148160	1.3	148859	1.3	149373	1.4	0.47
	중졸	107	-	47507	1.6	47614	1.6	78386	1.0	0.22
	고졸	2083	12.1	113560	1.6	115643	1.5	160486	1.1	1.80
	대졸이상	961	15.9	36483	1.9	37444	1.9	21993	3.3	2.57
계	18,422		821,847		840,269		649,531		2.19	

⑬ 전라남도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1553	19.1	157872	1.1	159425	1.1	58915	2.3	0.97
	중졸	2320	20.9	88379	1.2	90699	1.2	55077	1.5	2.56
	고졸	11228	7.2	256165	0.8	267393	0.7	87551	1.6	4.20
	대졸이상	2224	10.0	70721	0.6	72945	0.5	7478	3.6	3.05
여	초졸이하	1186	19.5	232345	1.5	233531	1.4	158687	1.6	0.51
	중졸	155	80.9	56889	1.7	57043	1.6	68124	1.0	0.27
	고졸	2260	13.3	115405	1.2	117665	1.3	121143	0.9	1.92
	대졸이상	664	7.6	23255	1.5	23919	1.4	10501	2.4	2.78
계	21,590		1,001,031		1,022,620		567,476		2.11	

⑭ 경상북도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1673	11.5	170233	0.8	171906	0.8	77092	1.4	0.97
	중졸	1799	5.1	113904	0.5	115703	0.5	72161	0.7	1.55
	고졸	13333	3.0	426761	0.4	440094	0.4	105288	1.3	3.03
	대졸이상	2400	3.2	109670	0.2	112070	0.2	9963	1.6	2.14
여	초졸이하	640	17.2	228061	1.1	228701	1.1	216678	0.9	0.28
	중졸	1345	7.5	84771	0.7	86115	0.7	94750	0.5	1.56
	고졸	4462	4.5	200024	0.7	204486	0.6	196756	0.5	2.18
	대졸이상	668	4.8	34271	0.6	34938	0.5	21740	0.7	1.91
계	26,320		1,367,695		1,394,013		794,428		1.89	

⑮ 경상남도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	1847	13.2	125367	1.0	127214	1.0	73453	1.7	1.45
	중졸	2636	8.3	110999	0.6	113635	0.6	91118	0.7	2.32
	고졸	20069	4.0	459398	0.4	479467	0.4	130755	1.4	4.19
	대졸이상	4091	5.6	133137	0.3	137228	0.2	14739	1.8	2.98
여	초졸이하	755	-	180629	1.1	181384	1.1	219911	0.9	0.42
	중졸	2169	17.6	88461	1.1	90630	1.1	121072	0.8	2.39
	고졸	7043	5.1	228610	0.8	235653	0.8	270186	0.7	2.99
	대졸이상	1214	3.5	43082	0.7	44296	0.7	33489	0.9	2.74
계	39,824		1,369,683		1,409,507		954,723		2.83	

(16) 제주도

구 분	실업자 수		취업자 수		경활인구		비경활인구		실업률 (%)	
	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)	추정값	CV (%)		
남	초졸이하	79	-	19830	0.2	19910	0.2	7892	0.3	0.40
	중졸	132	1.6	21964	0.2	22096	0.1	12156	0.2	0.60
	고졸	1119	0.5	96424	0.1	97543	0.1	19327	0.3	1.15
	대졸이상	242	0.6	30967	0.1	31208	0.1	2515	0.5	0.78
여	초졸이하	39	4.8	36264	0.3	36303	0.2	29447	0.2	0.11
	중졸	99	-	16013	0.2	16112	0.2	15830	0.2	0.61
	고졸	389	0.6	47553	0.2	47942	0.2	38326	0.2	0.81
	대졸이상	75	-	10252	0.2	10326	0.2	5432	0.3	0.73
계	2,174		279,267		281,440		130,925		0.77	

부록3 시군구 추정 SAS프로그램

```
/******  
    각 시도별로 취업자/실업자를 카운트하는 프로그램  
    (새로운 조사구 코드용)  
*****/  
OPTIONS ls=132 ps=200 CENTER NODATE NONUMBER  
FONT= 'Terminal' 8;  
  
/*=====*/  
%LET file_1 = 'd:\mySASpgm\0210\unemp\rawdata\m0210.txt'; /* 경찰인구 조사 결과  
원본 파일 */  
%LET file_2 = 'd:\mySASpgm\0210\unemp\rawdata\0210추계인구.xls';  
%LET lib_1 = project;  
  
/* data set의 라이브러리명 */  
%LET libpath_1 = 'd:\mySASpgm\0210\unemp\sasdata';/* 생성된dataset 저장 경로 */  
/*=====*/  
  
LIBNAME &lib_1 &libpath_1;  
  
DATA rawdata;  
    INFILE &file_1;  
    INPUT josagu 2-8 sex 21 age 22-24 educat 31-33 job 41 workable 50 ratio  
75-81 3;  
RUN;  
  
DATA raw;  
    SET rawdata;  
    IF job = 0 THEN jobtype = 1;  
        /* jobtype=1 : 취업자 */  
    ELSE IF job = 1 AND workable = 0 THEN jobtype = 1;  
    ELSE IF job = 1 AND workable = 1 THEN jobtype = 2;  
        /* jobtype=2 : 실업자 */  
    ELSE IF job = 2 THEN jobtype = 3;  
        /* jobtype=3 : 비경찰인구 */  
    ELSE DELETE;  
        /* 잘못입력된 경우 삭제 */  
    IF age < 15 THEN DELETE;  
    IF age < 30 THEN agetype=1;  
    ELSE agetype = 2;  
  
RUN;  
  
PROC SORT DATA = raw OUT = &lib_1.raw;
```

```

        BY josagu sex jobtype agetype;
RUN;

PROC FREQ DATA = &lib_1..raw;
    /* 조사구별 취업/실업/비경활 인구 카운트 */
    TABLES ratio josagu*sex*jobtype*agetype / OUT = counted1 NOPRINT
    SPARSE;
RUN;

DATA counted2;
    SET counted1;
    DROP percent;
RUN;

DATA ratiodata;
    /* 승수값 준비 */
    SET &lib_1..raw;
    BY josagu sex;
    IF FIRST.josagu = 1 OR FIRST.sex = 1;
    KEEP josagu sex ratio;
RUN;

PROC SQL;
    /* 조사구 | 성별 | 경제 범주 | 나이 범주 | 인원수 | 승수 */
    CREATE TABLE counted AS
    SELECT *
    FROM counted2 JOIN ratiodata ON counted2.josagu = ratiodata.josagu AND
        counted2.sex = ratiodata.sex;
    QUIT;
RUN;

%MACRO cutting(sheet, region);
    /* 광역시도별 데이터셋 분리 */
    DATA &lib_1..indat&region;
        SET counted;
        IF INT(josagu / 100000) = &region;
RUN;

PROC IMPORT
    /* 시도별 상주추정인구 파일 준비 */
    OUT= popul&region
    DATAFILE= &file_2
    DBMS=EXCEL2000 REPLACE;
    RANGE=&sheet;
RUN;

PROC TRANSPOSE

```

```

DATA = popul&region OUT = popul&region;
RUN;

DATA &lib_1..popul&region;
  SET popul&region;
  groupno = _N_ - 1;
  cityname = COMPRESS(_LABEL_, ' ');
  IF &region > 30 & SUBSTR(cityname, LENGTH(cityname)-1, 2) =
'군' THEN kind = 2;
  ELSE kind = 1;
  yboy = col13 ; oboy = col14 + col15 ;
  ygirl = col21 ; ogirl = col22 + col23 ;
  IF _N_ > 1;
  KEEP groupno cityname kind yboy oboy ygirl ogirl;
RUN ;
%MEND;

%cutting("대구$B5:AG28", 22); /* 대구광역시 */
%cutting("인천$B5:AG28", 23); /* 인천광역시 */
%cutting("광주$B5:AG28", 24); /* 광주광역시 */
%cutting("대전$B5:AG28", 25); /* 대전광역시 */
%cutting("울산$B5:AG28", 26); /* 울산광역시 */
%cutting("강원$B5:AG28", 32); /* 강원도 */
%cutting("충북$B5:AG28", 33); /* 충청북도 */
%cutting("충남$B5:AG28", 34); /* 충청남도 */
%cutting("전북$B5:AG28", 35); /* 전라북도 */
%cutting("전남$B5:AG28", 36); /* 전라남도 */
%cutting("경북$B5:AG28", 37); /* 경상북도 */
%cutting("경남$B5:AG28", 38); /* 경상남도 */
%cutting("제주$B5:AG28", 39); /* 제주도 */

```

```
/******
```

경제활동인구 계산 프로그램
(새로운 조사구 코드용)

```
*****/
```

```
OPTIONS ls=132 ps=200 CENTER NODATE NONUMBER  
FONT= 'Terminal' 8;
```

```
/*-----*/
```

```
%LET lib_1 = project;
```

```
/* dataset의 라이브러리명 */
```

```
%LET libpath_1 = 'd:\mySASpgm\0210\unemp\sasdata';
```

```
/* 생성된dataset 저장 경로 */
```

```
%LET lib_2 = result ;
```

```
/* 생성된 dataset을 저장할 라이브러리명 */
```

```
%LET libpath_2 = 'd:\mySASpgm\0210\unemp\result' ;
```

```
/* 생성된dataset의물리적경로 */
```

```
%LET lib_3 = report;
```

```
%LET libpath_3 = 'd:\mySASpgm\0210\unemp\report';
```

```
/*-----*/
```

```
LIBNAME &lib_1 &libpath_1 ;
```

```
LIBNAME &lib_2 &libpath_2 ;
```

```
LIBNAME &lib_3 &libpath_3 ;
```

```
/*-----
```

직접추정값 계산 루틴

```
-----*/
```

```
%MACRO process(josacode, outname, num);
```

```
DATA sigungroup;
```

```
INFILE &josacode;
```

```
INPUT groupno josagu;
```

```
RUN;
```

```
DATA groupname;
```

```
/* 소지역번호 | 시군 이름 | 소지역 종류 */
```

```
SET &lib_1..popul&num;
```

```
KEEP groupno cityname kind;
```

```
RUN;
```

```
DATA inputdata; /* 데이터 파일 */
```

```
SET &lib_1..indat&num ;
```

```
RUN ;
```



```

PROC SQL;
/* 소지역 번호 | 조사구 | 시군명 | 성별 | 경제 범주 | 나이 범주 | 인원수 | 승수 */
CREATE TABLE grouped AS
SELECT *
FROM sigungroup JOIN inputdata ON sigungroup.josagu =
inputdata.josagu;
QUIT;
RUN;

PROC SORT
/* 소지역별 소팅 */
DATA = grouped OUT = grouped;
BY groupno;
RUN;

DATA direct1;
SET grouped;
popul = count * ratio;
IF MOD(josagu, 100000) < 20000 THEN regiontype = 1;
ELSE regiontype = 2;
RUN;

PROC MEANS
/* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
DATA = direct1 NOPRINT NWAY;
OUTPUT OUT = direct2 SUM = popul;
CLASS groupno jobtype;
BY groupno;
VAR popul;
RUN;

DATA direct2;
/* 소지역번호 | 경제인 범주 | 직접추정치 */
SET direct2;
DROP josagu _TYPE_ _FREQ_;
RUN;

PROC IML WORKSPACE = 1024 SYMSIZE = 512;

CREATE outdir VAR {groupno estemp varemp estunemp varunemp
estnoact varnoact};

USE direct2 VAR {groupno jobtype popul};
/* 직접추정치 정보 */
READ ALL INTO estim;
USE direct1 VAR {groupno josagu sex jobtype agetype count ratio

```

```

popul regiontype); /* 분산 계산위해 */

      nofgroups = NROW(estim) / 3;
      DO i = 1 TO nofgroups;
          /* 소지역 갯수 만큼 반복 */
          DO s = 1 TO 2;
              /* 성별 */
              READ ALL WHERE (groupno = i & sex = s &
                             /* 소지역의 취업자 데이터 */
                             jobtype = 1) INTO emp;
              READ ALL WHERE (groupno = i & sex = s &
                             /* 소지역의 실업자 데이터 */
                             jobtype = 2) INTO unemp;
              READ ALL WHERE (groupno = i & sex = s &
                             /* 소지역의 비경활 데이터 */
                             jobtype = 3) INTO noact;
              IF s = 1 THEN DO;
                  nofjosagu = NROW(emp) / 2;
                  /* 소지역의 조사구 갯수 */
                  pivotrow = 0;
                  IF emp[1,2]/100000 > 30 & emp[1,9] ^=
emp[nofjosagu*2,9] THEN /* 시/읍면 혼합 조사구 */
                      DO j = nofjosagu*2 TO 2 BY -2;
                          IF emp[j,9] = 2 THEN pivotrow = j;
                      /* pivotrow : 읍면 조사구 시작 지점 */
                      END;
                      Zemp = SHAPE(0, nofjosagu, 2);
                      /* 취업자의 분산 계산을 위해 */
                      Zunemp = SHAPE(0, nofjosagu, 2);
                      /* 실업자의 분산 계산을 위해 */
                      Znoact = SHAPE(0, nofjosagu, 2);
                      /* 비경활의 분산 계산을 위해 */
                      END;
                      totE = SUM(emp[,6]);
                      /* 취업자 합계 */
                      totU = SUM(unemp[,6]);
                      /* 실업자 합계 */
                      totN = SUM(noact[,6]);
                      /* 비경활인구 합계 */
                      totX = totE + totU;
                      /* 경찰인구 합계 */
                      totT = totX + totN;
                      /* 조사인구 합계(경찰+비경찰) */

                      DO j = 1 TO nofjosagu - 1;
                          /* 조사구 수 만큼 반복 */
                          josa1 = emp[(j-1)*2+1, 2];
                          /* (j) 번째 조사구 번호 */
                          josa2 = emp[j*2+1, 2];
                          /* (j+1) 번째 조사구 번호 */

```

```

sex = s & jobtype = 1) INTO emp1;
sex = s & jobtype = 2) INTO unemp1;
sex = s & jobtype = 3) INTO noact1;
sex = s & jobtype = 1) INTO emp2;
sex = s & jobtype = 2) INTO unemp2;
sex = s & jobtype = 3) INTO noact2;

READ ALL WHERE (josagu = josal &
/* j 조사구의 취업자 데이터 */
READ ALL WHERE (josagu = josal &
/* j 조사구의 실업자 데이터 */
READ ALL WHERE (josagu = josal &
/* j 조사구의 비경활 데이터 */
READ ALL WHERE (josagu = josa2 &
/* j+1 조사구의 취업자 데이터 */
READ ALL WHERE (josagu = josa2 &
/* j+1 조사구의 실업자 데이터 */
READ ALL WHERE (josagu = josa2 &
/* j+1 조사구의 비경활 데이터 */

Ej1 = SUM(emp1[,6]);
/* j 조사구의 취업자 합계 */
Uj1 = SUM(unemp1[,6]);
/* j 조사구의 실업자 합계 */
Nj1 = SUM(noact1[,6]);
/* j 조사구의 비경활인구 합계 */
Xj1 = Ej1 + Uj1;
/* j 조사구의 경활인구 합계 */
Tj1 = Xj1 + Nj1;
/* j 조사구의 조사인구 합계 */

Ej2 = SUM(emp2[,6]);
/* j+1 조사구의 취업자 합계 */
Uj2 = SUM(unemp2[,6]);
/* j+1 조사구의 실업자 합계 */
Nj2 = SUM(noact2[,6]);
/* j+1 조사구의 비경활인구 합계 */
Xj2 = Ej2 + Uj2;
/* j+1 조사구의 경활인구 합계 */
Tj2 = Xj2 + Nj2;
/* j+1 조사구의 조사인구 합계 */

Zemp[j,s] = (Ej2 - Ej1) - ((Tj2 - Tj1) * totE / totT);
/* Zemp 계산 */
Zunemp[j,s] = (Uj2 - Uj1) - ((Tj2 - Tj1) * totU / totT);
/* Zunemp 계산 */
Znoact[j,s] = (Nj2 - Nj1) - ((Tj2 - Tj1) * totN / totT);
/* Znoact 계산 */
END;
/* 조사구 */
IF s = 1 THEN mrat1 = emp[1,7];
/* 남여 승수 */
ELSE frat1 = emp[1,7];

```

```

        IF pivotrow ^= 0 THEN DO;
        IF s = 1 THEN mrat2 = emp[pivotrow,7];
        ELSE frat2 = emp[pivotrow,7];
        Zemp[pivotrow/2-1, ] = 0;
        Zunemp[pivotrow/2-1, ] = 0;
        Znoact[pivotrow/2-1, ] = 0;
        END;
    END;
/* 성별 */
mCOVe = Zemp[,1] # Zemp[,2];
/* 취업자의 COV 계산(Zemp[0] * Zemp[1]) */
mCOVu = Zunemp[,1] # Zunemp[,2];
mCOVn = Znoact[,1] # Znoact[,2];

mtmpe = Zemp # Zemp;
/* 제공 */
mtmpu = Zunemp # Zunemp;
/* 제공 */
mtmpn = Znoact # Znoact;
/* 제공 */

IF pivotrow = 0 THEN nofjosagul = nofjosagu;
ELSE DO;
    nofjosagul = pivotrow / 2 - 1;
    /* 시지역의 조사구 갯수 */
    nofjosagu2 = nofjosagu - pivotrow / 2 + 1;
    /* 읍면의 조사구 갯수 */
    END;
mull = mrat1 * mrat1 * nofjosagul / (2*(nofjosagul-1));
/* 승수와 조사구수 고려 */
mul2 = frat1 * frat1 * nofjosagul / (2*(nofjosagul-1));
tmpel = SUM(mtmpe[,1]);
/* 남자 취업자 Zemp의 제공 합계 */
tmpe2 = SUM(mtmpe[,2]);
/* 여자 취업자 Zemp의 제공 합계 */
tmpul = SUM(mtmpu[,1]);
/* 남자 실업자 */
tmpu2 = SUM(mtmpu[,2]);
tmpn1 = SUM(mtmpn[,1]);
/* 남자 비경활인 */
tmpn2 = SUM(mtmpn[,2]);
IF pivotrow = 0 THEN DO;
    tmpel = tmpel * mull;
    /* Zemp 값에 위의 값 곱하기 */
    tmpe2 = tmpe2 * mul2;
    tmpul = tmpul * mull;
    tmpu2 = tmpu2 * mul2;

```

```

        tmpn1 = tmpn1 * mul1;
        tmpn2 = tmpn2 * mul2;
    END;
    ELSE DO;

        /* 혼합조사구의 경우(읍면 조사구 합계) */
        mul3 = mrat2 * mrat2 * nofjosagu2 /
        /* 승수와 조사구수 고려 */
        (2*(nofjosagu2-1));
        mul4 = frat2 * frat2 * nofjosagu2 /
        /* 승수와 조사구수 고려 */
        (2*(nofjosagu2-1));
        umpel = SUM(mtmpel[pivotrow/2-1:nofjosagu-1,1]);
        /* 남자 취업자 Zemp의 제공 합계 */
        umpe2 = SUM(mtmpel[pivotrow/2-1:nofjosagu-1,2]);
        /* 여자 취업자 Zemp의 제공 합계 */
        umpu1 = SUM(mtmpu[pivotrow/2-1:nofjosagu-1,1]);
        /* 남자 실업자 */
        umpu2 = SUM(mtmpu[pivotrow/2-1:nofjosagu-1,2]);
        umpn1 = SUM(mtmpn[pivotrow/2-1:nofjosagu-1,1]);
        /* 남자 비경활인 */
        umpn2 = SUM(mtmpn[pivotrow/2-1:nofjosagu-1,2]);
        tmpe1 = (tmpe1-umpel) * mul1 + umpel * mul3;
        tmpe2 = (tmpe2-umpe2) * mul2 + umpe2 * mul4;
        tmpu1 = (tmpu1-umpu1) * mul1 + umpu1 * mul3;
        tmpu2 = (tmpu2-umpu2) * mul2 + umpu2 * mul4;
        tmpn1 = (tmpn1-umpn1) * mul1 + umpn1 * mul3;
        tmpn2 = (tmpn2-umpn2) * mul2 + umpn2 * mul4;
    END;

    COVe = SUM(mCOVe);
    COVu = SUM(mCOVu);
    COVn = SUM(mCOVn);

    mul1 = mrat1 * frat1 * nofjosagu1 / (2*(nofjosagu1-1));
    /* 승수와 조사구수 고려 */
    IF pivotrow = 0 THEN DO;

        /* 혼합조사구의 경우 */
        COVe = COVe * mul1;

        /* COVe 에 위의 값 곱하기 */
        COVu = COVu * mul1;
        COVn = COVn * mul1;

    END;
    ELSE DO;
        mul2 = mrat2 * frat2 * nofjosagu2 / (2*(nofjosagu2-1));
        /* 승수와 조사구수 고려 */
        uCOVe = SUM(mCOVe[pivotrow/2-1:nofjosagu-1]);

```

```

uCOVu = SUM(mCOVu[pivotrow/2-1:nofjosagu-1]);
uCOVn = SUM(mCOVn[pivotrow/2-1:nofjosagu-1]);
COVe = (COVe-uCOVe) * mul1 + uCOVe * mul2;
/* COVe 에 위의 값 곱하기 */
COVu = (COVu-uCOVu) * mul1 + uCOVu * mul2;
COVn = (COVn-uCOVn) * mul1 + uCOVn * mul2;
END;

groupno = i;
varemp = SQRT(tmpe1 + tmpe2 + 2*COVe);
varunemp = SQRT(tmpu1 + tmpu2 + 2*COVu);
varnoact = SQRT(tmpn1 + tmpn2 + 2*COVn);
estemp = estim[(i-1)*3+1, 3];
estunemp = estim[(i-1)*3+2, 3];
/*
*/
IF estunemp = 0 THEN estunemp = mrat1;
/* 실업자수가 0일때 보정 */
estnoact = estim[(i-1)*3+3, 3];
APPEND;
/* 취업자 추정값 및 분산, 실업자 추정값 및 분산, 비경활인 추정값 및 분산 출력 */
END;
/* 소지역 */
CLOSE outdir;
QUIT;

PROC SQL;
/* 직접추정값 출력 */
CREATE TABLE &lib_2..outdir AS
/* 지역번호 | 시군 이름 | 지역종류(시/읍면) | 취업자 추정값 | 분산 | 실업자 추정
값 | 분산 | 비경활 추정값 | 분산 */
SELECT *
FROM groupname JOIN outdir ON groupname.groupno =
outdir.groupno;
QUIT;

/*=====
합성 및 복합추정값(재크나이프) 계산 루틴
=====*/

PROC MEANS
/* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
DATA = direct1 NOPRINT NWAY;
OUTPUT OUT = direct2 SUM = count popul;
CLASS groupno sex jobtype agetype;
BY groupno;
VAR count popul;

```

```

RUN;

DATA direct2; /* 소지역번호 | 경제인 범주 | 직접추정치 */
    SET direct2;
    DROP josagu _TYPE_ _FREQ_;
RUN;

PROC IML WORKSPACE = 1024 SYMSIZE = 512;

START jknife(from, to, jcode);
/* 재크나이프 모듈 선언 : 시/읍면 단위 조사구 각각 처리하기 위함 */
    USE &lib_1..popul&num VAR {kind yboy oboy ygirl ogirl};
    /* 상주추정인구 파일 */
    READ ALL VAR {yboy oboy ygirl ogirl} WHERE (kind = jcode)
INTO A; /* A : 상주추정인구 */
    USE direct2 VAR {groupno sex jobtype agetype count};
    /* 경찰조사 데이터(소지역별) */
    READ ALL INTO estim;
    USE &lib_2..outdir VAR {groupno kind estunemp varunemp};
    /* 직접추정치 및 분산 입력 */
    READ ALL WHERE (kind = jcode) INTO dirval;

    ngrps = to - from + 1; /* 소지역의 수 */
    x = SHAPE(0, ngrps, 4); /* 범주별 취업자수 */
    y = SHAPE(0, ngrps, 4); /* 범주별 실업자수 */
    z = SHAPE(0, ngrps, 4); /* 범주별 비경활인수 */
    dde = SHAPE(0, 4, 1);
    ddu = SHAPE(0, 4, 1);
    ddn = SHAPE(0, 4, 1);
    Ys = SHAPE(0, ngrps, 1);
    Yc = SHAPE(0, ngrps, 1);

    DO i = from TO to;
        x[i-from+1,1] = estim[(i-1)*12+1, 5];
x[i-from+1,2] = estim[(i-1)*12+2, 5];
        x[i-from+1,3] = estim[(i-1)*12+7, 5];
x[i-from+1,4] = estim[(i-1)*12+8, 5];
        y[i-from+1,1] = estim[(i-1)*12+3, 5];
y[i-from+1,2] = estim[(i-1)*12+4, 5];
        y[i-from+1,3] = estim[(i-1)*12+9, 5];
y[i-from+1,4] = estim[(i-1)*12+10, 5];
        z[i-from+1,1] = estim[(i-1)*12+5, 5];
z[i-from+1,2] = estim[(i-1)*12+6, 5];
        z[i-from+1,3] = estim[(i-1)*12+11, 5];
= estim[(i-1)*12+12, 5];
        z[i-from+1,4]
    END;

```

```

C = y;          /* C : 범주별 실업자수 */
D = x + y;     /* D : 범주별 경찰인구 */
B = D + z;     /* B : 범주별 조사인구 */

x = A / B # C;
Ydir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) || SUM(x[,4]);
/* Ydir : 범주별 직접추정치 */
tA = SUM(A[,1]) || SUM(A[,2]) || SUM(A[,3]) || SUM(A[,4]);
/* A의 범주별 합계 */
DO i = 1 TO ngrps;

                /* Ys : 합성추정값 */
                zYs = (A[i,] # Ydir) / tA;
                Ys[i] = SUM(zYs);
END;
USE direct1 VAR {groupno josagu sex jobtype agetype count};
/* 경찰조사 데이터(조사구별) */
SETOUT outsyncom;

DO i = from TO to; /* 소지역 갯수 만큼 반복 */
emp;
unemp;
noact;
    READ ALL WHERE (groupno = i & jobtype = 1) INTO
    /* 조사구별 취업자 데이터 */
    READ ALL WHERE (groupno = i & jobtype = 2) INTO
    /* 조사구별 실업자 데이터 */
    READ ALL WHERE (groupno = i & jobtype = 3) INTO
    /* 조사구별 비경찰 데이터 */
    nofjosagu = NROW(emp) / 4;
    iYs = SHAPE(0, nofjosagu, 1);
    DO without = 1 TO nofjosagu;
    /* 조사구 1개씩 제외하면서 B, C, D 재계산 */
        iB = B;          iC = C;          iD = D;
    /* 제외하지 않은 B, C, D값 저장 */
        DO j = 1 TO 4;
        /* 연령 범주 */
            dde[j] = emp[(without-1)*4+j, 6];
            ddu[j] = unemp[(without-1)*4+j, 6];
            ddn[j] = noact[(without-1)*4+j, 6];
        END;
        iemp = dde`; iunemp = ddu`; inoact = ddn` ;
    /* Transpose */
    B[i-from+1,] = B[i-from+1,] - (iemp + iunemp +
    /* 조사구 1개 제외된 B, C, D */
    C[i-from+1,] = C[i-from+1,] - iunemp;

    x = A / B # C;
    iYdir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) ||

```



```

SUM(x[,4]);          /* iYdir : 조사구 제외후 범주별 직접추정치 */
                    zYs = (A[i-from+1,] # iYdir) / tA;
                    iYs[without] = SUM(zYs);
                    B[i-from+1,] = B[i-from+1,] + (iemp + iunemp +
inoact);          /* 제외 이전으로 B, C, D 복구 */
                    C[i-from+1,] = C[i-from+1,] + iunemp;
END;
/* 합성추정값 계산 시작 */
avgYs = SUM(iYs) / nofjosagu;
varYs = iYs - avgYs;
varYs = varYs # varYs;
varYs = SUM(varYs) * (nofjosagu-1) / nofjosagu;
groupno = i;
synest = Ys[i-from+1];
synbias = (nofjosagu-1) * (avgYs - Ys[i-from+1]);
IF varYs < 0.0001 THEN DO;
/* 연산 오차 고려 */
                    varYs = 0;
                    synbias = 0;
END;
synvar = varYs;
mse = varYs + synbias * synbias;
synmse = SQRT(mse);
syncv = synmse / Ys[i-from+1] * 100;
IF syncv = . THEN syncv = 0;
/* 합성추정값 계산 완료 */
/* 복합추정값 계산 시작 */
direst = dirval[i-from+1, 3];
dirvar = dirval[i-from+1, 4];
dirvar = dirvar ** 2;
/* 직접추정치의 분산 */
IF (mse + dirvar) = 0 THEN wgt = 0;
ELSE wgt = mse / (mse + dirvar);
/* Weight 계산 */
Yc[i-from+1] = wgt * direst + (1-wgt) * Ys[i-from+1];
iYc = wgt * direst + (1-wgt) * iYs;
avgYc = SUM(iYc) / nofjosagu;
varYc = iYc - avgYc;
varYc = varYc # varYc;
varYc = SUM(varYc) * (nofjosagu-1) / nofjosagu;
comest = Yc[i-from+1];
combias = (nofjosagu - 1) * (avgYc - Yc[i-from+1]);
IF varYc < 0.0001 THEN DO;
/* 연산 오차 고려 */
                    varYc = 0;
                    combias = 0;
END;

```

```

        comvar = varYc;
        commse = SQRT(varYc + combias * combias);
        comcv = commse / Yc[i-from+1] * 100;
        IF comcv = . THEN comcv = 0;
        /* 복합추정값 계산 완료 */
        APPEND;
    END;
FINISH;

CREATE outsyncom VAR {groupno synest synbias synvar synmse
syncv comest combias comvar commse comcv};
USE groupname VAR {groupno kind};
READ ALL INTO grpname;

USE &lib_2..outdir VAR {groupno kind estunemp varunemp};
        /* 직접추정치 및 분산 입력 */
READ ALL INTO dirval;

nofgroups = NROW(dirval);

splitrow = 0;

                                /* splitrow : 군지역 시작위치 */
DO i = nofgroups TO 1 BY -1;
    IF grpname[i,2] = 2 THEN splitrow = i;
END;
IF splitrow = 0 THEN RUN jknife(1, nofgroups, 1);
        /* 시지역만 계산 */
ELSE DO;
    IF splitrow ^= 0 THEN RUN jknife(1, splitrow-1, 1);
        /* 시지역 계산 */
    IF splitrow ^= 0 THEN RUN jknife(splitrow, nofgroups, 2);
    /* 군지역 계산 */
END;

CLOSE outsyncom;
QUIT;

DATA &lib_2..outsyncom;
    SET outsyncom;
RUN;

DATA dir2;
    SET &lib_2..outdir;
    IF estunemp = 0 THEN dircv = 0;
    ELSE dircv = varunemp / estunemp * 100;
    KEEP groupno kind estunemp varunemp dircv;

```

```

RUN;

PROC SQL;
                                /* 계산 결과 테이블 조인 */
    CREATE TABLE tmpresult AS
    SELECT *
    FROM dir2 JOIN outsyncom ON dir2.groupno = outsyncom.groupno;
QUIT;

PROC IML;    /* 직접추정치가 0인 경우 Common Weight 적용*/
    CREATE corcom VAR (groupno comest);
    USE tmpresult VAR {groupno kind estunemp synest synmse
comest};

    READ ALL INTO tmptable;
    pivot = 0;                    /* 시지역 / 군지역 구분 */
    theend = NROW(tmptable);
    DO i = 1 TO theend;
        IF tmptable[i, 2] = 1 THEN pivot = pivot + 1;
    END;
    DO i = 1 TO theend;
        groupno = tmptable[i, 1];
        IF tmptable[i, 3] > 0 THEN
            comest = tmptable[i, 6];
        ELSE DO;
            msej = tmptable[i, 5] # tmptable[i, 5];
            subY = tmptable[i, 4] - tmptable[i, 3];
            subY = subY # subY;
            IF i <= pivot THEN
                wgt = SUM(msej[1:pivot]) / SUM(subY[1:pivot]);
            ELSE
                wgt = SUM(msej[pivot+1:theend]) / SUM(subY[pivot+1:theend]);
            comest = (1 - wgt) * tmptable[i, 4];
        END;
        APPEND;
    END;
    CLOSE corcom;
QUIT;

DATA tmpresult;
    SET tmpresult;
    DROP comest;
RUN;

PROC SQL;
    CREATE TABLE res AS
    SELECT *
    FROM tmpresult JOIN corcom ON tmpresult.groupno = corcom.groupno;

```

```

QUIT;

PROC IML;
                                /* 보정값 계산 */
    CREATE revval VAR {groupno revsyn revcom};
    USE res VAR {groupno kind estunemp synest comest};
    READ ALL INTO restable;
    pivot = 0;                    /* 시지역 / 군지역 구분 */
    theend = NROW(restable);
    DO i = 1 TO theend;
        IF restable[i, 2] = 1 THEN pivot = pivot + 1;
    END;
    DO i = 1 TO theend;
        groupno = restable[i, 1];
        IF i <= pivot THEN DO;
            revsyn = SUM(restable[1:pivot, 3]) * restable[i, 4] / SUM(restable[1:pivot, 4]);
            revcom = SUM(restable[1:pivot, 3]) * restable[i, 5] / SUM(restable[1:pivot, 5]);
        END;
        ELSE DO;
            revsyn = SUM(restable[pivot+1:theend, 3]) *
restable[i, 4] / SUM(restable[pivot+1:theend, 4]);
            revcom = SUM(restable[pivot+1:theend, 3]) *
restable[i, 5] / SUM(restable[pivot+1:theend, 5]);
        END;
        IF revsyn = . THEN revsyn = 0;
        IF revcom = . THEN revcom = 0;
        APPEND;
    END;
    CLOSE revval;

QUIT;

DATA res;
    SET res;
    DROP kind;

RUN;

PROC SQL;
    CREATE TABLE result1 AS
    SELECT *
    FROM res JOIN revval ON res.groupno = revval.groupno;

QUIT;

DATA groupname;
    SET groupname;
    DROP kind;

RUN;

PROC SQL;

```

```

CREATE TABLE result AS
SELECT *
FROM groupname JOIN result1 ON groupname.groupno =
result1.groupno;
QUIT;

DATA &outname;
SET result;
DROP groupno;
LABEL cityname = '소지역이름'
      estunemp = '실업자수[직접]'
      varunemp = '추정오차'
      dircv = 'CV[직접]'
      synest = '실업자수[합성]'
      synbias = 'Bias'
      synvar = '분 산'
      symmse = 'root.mse'
      syncv = 'CV[합성]'
      comest = '실업자수[복합]'
      combias = ' Bias   [복합]'
      comvar = '분 산'
      commse = 'root.mse'
      comcv = 'CV'
      revsyn = '보 정 값[합성]'
      revcom = '보 정 값[복합]';

RUN;
%MEND;

```

```

/* 특별시 및 광역시 처리 */
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\대구조사구.txt",
"d:\mySASpgm\0210\unemp\report\Taegu", 22);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\인천조사구.txt",
"d:\mySASpgm\0210\unemp\report\Incheon", 23);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\광주조사구.txt",
"d:\mySASpgm\0210\unemp\report\Gwangju", 24);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\대전조사구.txt",
"d:\mySASpgm\0210\unemp\report\Daejon", 25);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\울산조사구.txt",
"d:\mySASpgm\0210\unemp\report\Ulsan", 26);

```

```

/* 도 처리 */
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\강원조사구.txt",
"d:\mySASpgm\0210\unemp\report\Kangwon", 32);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\충북조사구.txt",
"d:\mySASpgm\0210\unemp\report\Chbuk", 33);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\충남조사구.txt",
"d:\mySASpgm\0210\unemp\report\Chnam", 34);

```

```

%process("d:\mySASpgm\0210\unemp\rawdata\josacode\전북조사구.txt",
"d:\mySASpgm\0210\unemp\report\Jeonbuk", 35);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\전남조사구.txt",
"d:\mySASpgm\0210\unemp\report\Jeonnam", 36);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\경북조사구.txt",
"d:\mySASpgm\0210\unemp\report\Kyongbuk", 37);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\경남조사구.txt",
"d:\mySASpgm\0210\unemp\report\Kyongnam", 38);
%process("d:\mySASpgm\0210\unemp\rawdata\josacode\제주조사구.txt",
"d:\mySASpgm\0210\unemp\report\Jeju", 39);

```

/* 결과 출력 */

```

PROC PRINT DATA = &lib_3.Taegu LABEL;
    TITLE '대구광역시 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Incheon LABEL;
    TITLE '인천광역시 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Gwangju LABEL;
    TITLE '광주광역시 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Daejeon LABEL;
    TITLE '대전광역시 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Ulsan LABEL;
    TITLE '울산광역시 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Kangwon LABEL;
    TITLE '강원도 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Chbuk LABEL;
    TITLE '충청북도 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Chnam LABEL;
    TITLE '충청남도 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Jeonbuk LABEL;
    TITLE '전라북도 실업자 현황';
RUN;

```

```
PROC PRINT DATA = &lib_3.Jeonnam LABEL;  
    TITLE '전라남도 실업자 현황';  
RUN;
```

```
PROC PRINT DATA = &lib_3.Kyongbuk LABEL;  
    TITLE '경상북도 실업자 현황';  
RUN;
```

```
PROC PRINT DATA = &lib_3.Kyongnam LABEL;  
    TITLE '경상남도 실업자 현황';  
RUN;
```

```
PROC PRINT DATA = &lib_3.Jeju LABEL;  
    TITLE '제주도 실업자 현황';  
RUN;
```

부록4 시·도단위 세부영역 추정 SAS 프로그램

(1) 성별-연령대별 추정 SAS 프로그램

```
/******  
  각 시도별로 취업자/실업자를 카운트하는 프로그램  
  (연령대별/성별, 새로운 조사구 코드용)  
******/  
  
OPTIONS ls=132 ps=200 CENTER NODATE NONUMBER  
FONT= 'Terminal' 8;  
  
/*=====*/  
%LET file_1 = 'd:\mySASpgm\0210\age\unemp\rawdata\m0210.txt';  
  /* 경찰인구 조사 결과 원본 파일 */  
%LET file_2 = 'd:\mySASpgm\0210\age\unemp\rawdata\0210연령별상주인구.xls';  
  /* 상주인구 데이터 */  
%LET lib_1 = project;  
  
  /* data set의 라이브러리명 */  
%LET libpath_1 = 'd:\mySASpgm\0210\age\unemp\sasdata';  
  /* 생성된dataset 저장 경로 */  
/*=====*/  
  
LIBNAME &lib_1 &libpath_1;  
  
DATA &lib_1..raw;  
  INFILE &file_1;  
  INPUT josagu 2-8 sex 21 age 22-24 edulevel 31 edustate 33 job 41 workable  
50 ratio 75-81 .3;  
RUN;  
  
DATA rawdata;  
  SET &lib_1..raw;  
  IF job = 0 THEN jobtype = 1;  
    /* jobtype=1 : 취업자 */  
  ELSE IF job = 1 AND workable = 0 THEN jobtype = 1;  
  ELSE IF job = 1 AND workable = 1 THEN jobtype = 2;  
    /* jobtype=2 : 실업자 */  
  ELSE IF job = 2 THEN jobtype = 3;  
    /* jobtype=3 : 비경찰인구 */  
  ELSE DELETE;  
    /* 잘못입력된 경우 삭제 */  
  
  IF age < 15 THEN DELETE;
```



```

IF age >= 50 THEN agetype = 4;
ELSE IF 40 <= age <= 49 THEN agetype = 3;
ELSE IF 30 <= age <= 39 THEN agetype = 2;
ELSE IF 15 <= age <= 29 THEN agetype = 1;

IF edulevel > 4 THEN edulevel = edulevel - 1;
/* 전문대 이상일 경우 */
IF edustate > 1 THEN edulevel = edulevel - 1;
/* 재학/중퇴/휴학의 경우 */
IF edulevel < 1 THEN edulevel = 1;
/* 초등졸 미만일 경우 */
IF edulevel > 4 THEN edulevel = 4;
/* 대졸 초과일 경우 */

RUN;

%MACRO cutting(sheet, pagetype, psex); /* 연령대별/성별 데이터셋 작성 */
DATA rawdata1;
SET rawdata;
IF sex = &psex AND agetype = &pagetype;
DROP sex agetype;
RUN;

PROC FREQ DATA = rawdata1;
/* 조사구별 취업/실업/비경활 인구 카운트 */
TABLES ratio josagu*jobtype*edulevel / OUT = counted1
NOPRINT SPARSE;
RUN;

DATA counted2;
SET counted1;
DROP percent;
RUN;

DATA ratiodata;
/* 승수값 준비 */
SET rawdata1;
BY josagu;
IF FIRST.josagu = 1;
KEEP josagu ratio;
RUN;

PROC SQL;
/* 조사구 | 경제 범주 | 인원수 | 승수 */
CREATE TABLE indat&pagetype&psex AS
SELECT *
FROM counted2 JOIN ratiodata ON counted2.josagu =

```

```

ratiodata.josagu;
    QUIT;
RUN;

PROC IMPORT
    /* 연령대별/성별 상주추정인구 파일 준비 */
    OUT= popul&pagetype&psex
DATAFILE= &file_2
    DBMS=EXCEL2000 REPLACE;
    RANGE=&sheet;
RUN;

DATA popul&pagetype&psex;
    SET popul&pagetype&psex;
    name = F1;
    jcode = _COL1;
    ctgry1 = _COL2;
    ctgry2 = _COL3;
    ctgry3 = _COL4;
    ctgry4 = _COL5;
    KEEP name jcode ctgry1 ctgry2 ctgry3 ctgry4;
RUN;

DATA &lib_1..popul1&pagetype&psex;
    SET popul&pagetype&psex;
    IF jcode < 30;
RUN;

DATA &lib_1..indat1&pagetype&psex;
    SET indat&pagetype&psex;
    jcode = INT(josagu / 100000);
    IF jcode < 30;

    /* 1집락 : 특별시 및 광역시 */
RUN;

DATA &lib_1..popul2&pagetype&psex;
    SET popul&pagetype&psex;
    IF jcode = 31 OR jcode = 34 OR 37 <= jcode <= 39;
RUN;

DATA &lib_1..indat2&pagetype&psex;
    SET indat&pagetype&psex;
    jcode = INT(josagu / 100000);
    IF jcode = 31 OR jcode = 34 OR 37 <= jcode <= 39;
    /* 2집락 : 경기, 충남, 경북, 경남, 제주 */
RUN;

```

```

DATA &lib_1..popul3&pagetype&psex;
  SET popul&pagetype&psex;
  IF 32 <= jcode <= 33 OR 35 <= jcode <= 36;
RUN;

DATA &lib_1..indat3&pagetype&psex;
  SET indat&pagetype&psex;
  jcode = INT(josagu / 100000);
  IF 32 <= jcode <= 33 OR 35 <= jcode <= 36;
  /* 3집락 : 강원, 충북, 전북, 전남 */
RUN;
%MEND;

%cutting("15-29남$A1:F17", 1, 1);          /* 15-29세 남자 */
%cutting("15-29여$A1:F17", 1, 2);        /* 15-29세 여자 */
%cutting("30-39남$A1:F17", 2, 1);        /* 30-39세 남자 */
%cutting("30-39여$A1:F17", 2, 2);        /* 30-39세 여자 */
%cutting("40-49남$A1:F17", 3, 1);        /* 40-49세 남자 */
%cutting("40-49여$A1:F17", 3, 2);        /* 40-49세 여자 */
%cutting("50+남$A1:F17", 4, 1);          /* 50세 이상 남자 */
%cutting("50+여$A1:F17", 4, 2);          /* 50세 이상 여자 */

```

```
/******
```

```
경제활동인구 계산 프로그램(연령대별/성별)
```

```
******/
```

```
OPTIONS ls=132 ps=200 CENTER NODATE NONUMBER  
FONT= 'Terminal' 8;
```

```
/*=====*/
```

```
%LET lib_1 = project;
```

```
/* dataset의 라이브러리명 */
```

```
%LET libpath_1 = 'd:\mySASpgm\0210\age\unemp\sasdata';
```

```
/* 생성된dataset 저장 경로 */
```

```
%LET lib_2 = result ;
```

```
/* 생성된 dataset을 저장할 라이브러리명 */
```

```
%LET libpath_2 = 'd:\mySASpgm\0210\age\unemp\result' ;
```

```
/* 생성된dataset의물리적경로 */
```

```
%LET lib_3 = report;
```

```
%LET libpath_3 = 'd:\mySASpgm\0210\age\unemp\report';
```

```
/*=====*/
```

```
LIBNAME &lib_1 &libpath_1 ;
```

```
LIBNAME &lib_2 &libpath_2 ;
```

```
LIBNAME &lib_3 &libpath_3 ;
```

```
/*=====
```

```
직접추정값 계산 루틴
```

```
=====*/
```

```
%MACRO process(category, outname);
```

```
DATA groupname; /* 시도이름 | 조사구 코드 */
```

```
SET &lib_1..popul&category;
```

```
KEEP name jcode;
```

```
RUN;
```

```
DATA inputdata; /* 데이터 파일 */
```

```
SET &lib_1..indat&category;
```

```
RUN ;
```

```
DATA direct1;
```

```
SET inputdata;
```

```
popul = count * ratio;
```

```
RUN;
```

```
PROC MEANS
```

```

/* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
DATA = direct1 NOPRINT NWAY;
OUTPUT OUT = direct2 SUM = popul;
CLASS jcode jobtype;
BY jcode;
VAR popul;

RUN;

DATA direct2;
/* 소지역번호 | 경제인 범주 | 직접추정치 */
SET direct2;
DROP josagu _TYPE_ _FREQ_;

RUN;

PROC IML SYMSIZE=1024;
CREATE outdir VAR {jcode estemp varemp estunemp varunemp
estnoact varnoact};

USE direct2 VAR {jcode jobtype popul};
/* 직접추정치 정보 */
READ ALL INTO estim;
USE direct1 VAR {josagu jobtype edulevel count ratio jcode popul};
/* 분산 계산위해 */

nofgroups = NROW(estim) / 3;
DO i = 1 TO nofgroups;
/* 소지역 갯수 만큼 반복 */
jcode2 = estim[i*3, 1];
emp; /* 소지역의 취업자 데이터 */
READ ALL WHERE (jcode = jcode2 & jobtype = 1) INTO
unemp; /* 소지역의 실업자 데이터 */
READ ALL WHERE (jcode = jcode2 & jobtype = 2) INTO
noact; /* 소지역의 비경활 데이터 */

nofjosagu = NROW(emp) / 4;
/* 소지역의 조사구 갯수 */
pivotrow = 0;
IF emp[1,1]/100000 > 30 & MOD(emp[nofjosagu*4,1],
100000) >= 20000 THEN /* 시군 혼합 지역 (도) */
DO j = nofjosagu*4 TO 4 BY -4;
IF MOD(emp[j,1], 100000) >= 20000 THEN pivotrow = j;
/* pivotrow : 읍면 조사구 시작 지점 */
END;
Zemp = SHAPE(0, nofjosagu, 1);
/* 취업자의 분산 계산을 위해 */
Zunemp = SHAPE(0, nofjosagu, 1);

```

```

/* 실업자의 분산 계산을 위해 */
      Znoact = SHAPE(0, nofjosagu, 1);
/* 비경활의 분산 계산을 위해 */

      totE = SUM(emp[,4]);
/* 취업자 합계 */
      totU = SUM(unemp[,4]);
/* 실업자 합계 */
      totN = SUM(noact[,4]);
/* 비경활인구 합계 */
      totX = totE + totU;
/* 경활인구 합계 */
      totT = totX + totN;
/* 조사인구 합계(경활+비경활) */

      DO j = 1 TO nofjosagu - 1;
/* 조사구 수 만큼 반복 */
          josag1 = emp[(j-1)*4+1, 1];
/* (j) 번째 조사구 번호 */
          josag2 = emp[j*4+1, 1];
/* (j+1) 번째 조사구 번호 */
          READ ALL WHERE (josagu = josag1 & jobtype
= 1) INTO emp1;
/* j 조사구의 취업자 데이터 */
          READ ALL WHERE (josagu = josag1 & jobtype
= 2) INTO unemp1;
/* j 조사구의 실업자 데이터 */
          READ ALL WHERE (josagu = josag1 & jobtype
= 3) INTO noact1;
/* j 조사구의 비경활 데이터 */
          READ ALL WHERE (josagu = josag2 & jobtype
= 1) INTO emp2;
/* j+1 조사구의 취업자 데이터 */
          READ ALL WHERE (josagu = josag2 & jobtype
= 2) INTO unemp2;
/* j+1 조사구의 실업자 데이터 */
          READ ALL WHERE (josagu = josag2 & jobtype
= 3) INTO noact2;
/* j+1 조사구의 비경활 데이터 */

          Ej1 = SUM(emp1[,4]);
/* j 조사구의 취업자 합계 */
          Uj1 = SUM(unemp1[,4]);
/* j 조사구의 실업자 합계 */
          Nj1 = SUM(noact1[,4]);
/* j 조사구의 비경활인구 합계 */
          Xj1 = Ej1 + Uj1;
/* j 조사구의 경활인구 합계 */
          Tj1 = Xj1 + Nj1;
/* j 조사구의 조사인구 합계 */

          Ej2 = SUM(emp2[,4]);
/* j+1 조사구의 취업자 합계 */

```

```

        Uj2 = SUM(unemp2[,4]);
/* j+1 조사구의 실업자 합계 */
        Nj2 = SUM(noact2[,4]);
/* j+1 조사구의 비경활인구 합계 */
        Xj2 = Ej2 + Uj2;
/* j+1 조사구의 경활인구 합계 */
        Tj2 = Xj2 + Nj2;
/* j+1 조사구의 조사인구 합계 */

        Zemp[j] = (Ej2 - Ej1) - ((Tj2 - Tj1) * totE / totT);
/* Zemp 계산 */
        Zunemp[j] = (Uj2 - Uj1) - ((Tj2 - Tj1) * totU / totT);
/* Zunemp 계산 */
        Znoact[j] = (Nj2 - Nj1) - ((Tj2 - Tj1) * totN / totT);
/* Znoact 계산 */
    END;                                /* 조사구 */

    rat1 = emp[1,5];
/* 시지역 승수 */
    IF pivotrow ^= 0 THEN DO;
        rat2 = emp[pivotrow,5];
/* 군지역 승수 */
        Zemp[pivotrow/4-1] = 0;
/* 접경지역의 Z 값 cut-off */
        Zunemp[pivotrow/4-1] = 0;
        Znoact[pivotrow/4-1] = 0;
    END;

    mtmpe = Zemp # Zemp;                /* 제공 */
    mtmpu = Zunemp # Zunemp;            /* 제공 */
    mtmpn = Znoact # Znoact;            /* 제공 */

    IF pivotrow = 0 THEN nofjosagu1 = nofjosagu;
    ELSE DO;
        nofjosagu1 = pivotrow / 4 - 1;
/* 시지역의 조사구 갯수 */
        nofjosagu2 = nofjosagu - pivotrow / 4 + 1;
/* 읍면의 조사구 갯수 */
    END;

    mull = rat1 * rat1 * nofjosagu1 / (2*(nofjosagu1-1));
/* 승수와 조사구수 고려 */
    tmpe = SUM(mtmpe);
/* 남자 취업자 Zemp의 제공 합계 */
    tmpu = SUM(mtmpu);
/* 남자 실업자 */
    tmpn = SUM(mtmpn);

```

```

/* 남자 비경활인 */
IF pivotrow = 0 THEN DO;
    tmpe = tmpe * mul1;
/* Zemp 값에 위의 값 곱하기 */
    tmpu = tmpu * mul1;
    tmpn = tmpn * mul1;
END;
ELSE DO;

/* 혼합조사구의 경우(읍면 조사구 합계) */
    mul2 = rat2 * rat2 * nofjosagu2 /
(2*(nofjosagu2-1));
/* 승수와 조사구수 고려 */
    umpe = SUM(mtmpe[pivotrow/4-1:nofjosagu-1]);
/* 남자 취업자 Zemp의 제곱 합계 */
    umpu = SUM(mtmpu[pivotrow/4-1:nofjosagu-1]);
/* 남자 실업자 */
    umpn = SUM(mtmpn[pivotrow/4-1:nofjosagu-1]);
/* 남자 비경활인 */
    tmpe = (tmpe-umpe) * mul1 + umpe * mul2;
    tmpu = (tmpu-umpu) * mul1 + umpu * mul2;
    tmpn = (tmpn-umpn) * mul1 + umpn * mul2;
END;

jcode = jcode2;
varemp = SQRT(tmpe);
varunemp = SQRT(tmpu);
varnoact = SQRT(tmpn);
estemp = estim[(i-1)*3+1, 3];
estunemp = estim[(i-1)*3+2, 3];
estnoact = estim[(i-1)*3+3, 3];
APPEND;
/* 취업자 추정값 및 분산, 실업자 추정값 및 분산, 비경활인 추정값 및 분산 출력 */
END;
/* 소지역 */
CLOSE outdir;
QUIT;

PROC SQL;
/* 직접추정값 출력 */
CREATE TABLE &lib_2.outdir AS
/* 시군 이름 | 조사구 코드 | 취업자 추정값 | 분산 | 실업자 추정값 | 분산 | 비경
활 추정값 | 분산 */
SELECT *
FROM groupname JOIN outdir ON groupname.jcode = outdir.jcode;
QUIT;

```



```

/*=====
합성 및 복합추정값(제크나이프) 계산 루틴
=====*/

```

```
PROC MEANS
```

```

/* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
DATA = direct1 NOPRINT NWAY;
OUTPUT OUT = direct2 SUM = count popul;
CLASS jcode jobtype edulevel;
BY jcode;
VAR count popul;

```

```
RUN;
```

```
DATA direct2;
```

```

/* 소지역번호 | 경제인 범주 | 직접추정치 */
SET direct2;
DROP josagu _TYPE_ _FREQ_;

```

```
RUN;
```

```
PROC IML;
```

```

CREATE outsyncom VAR {jcode synest synbias synvar symmse
syncv comest combias comvar commse comcv};
CREATE revised VAR {jcode revsyn revcom};

```

```

ctgry4);
USE &lib_1..popul&category VAR {jcode ctgry1 ctgry2 ctgry3
/* 상주추정인구 파일 */
READ ALL VAR {ctgry1 ctgry2 ctgry3 ctgry4} INTO A;
/* A : 상주추정인구 */
USE direct2 VAR {jcode jobtype count};

```

```

/* 경찰조사 데이터(소지역별) */
READ ALL INTO estim;
USE &lib_2..outdir VAR {jcode estunemp varunemp};
/* 직접추정치 및 분산 입력 */
READ ALL INTO dirval;

```

```

ngrps = NROW(dirval);
x = SHAPE(0, ngrps, 4);
y = SHAPE(0, ngrps, 4);
z = SHAPE(0, ngrps, 4);
dde = SHAPE(0, 4, 1);
ddu = SHAPE(0, 4, 1);
ddn = SHAPE(0, 4, 1);
Ys = SHAPE(0, ngrps, 1);
Yc = SHAPE(0, ngrps, 1);
/* 시도 갯수 */
/* 범주별 취업자수 */
/* 범주별 실업자수 */
/* 범주별 비경활인수 */

```

```

DO i = 1 TO ngrps;
  x[i,1] = estim[(i-1)*12+1, 3]; x[i,2] = estim[(i-1)*12+2, 3];
  x[i,3] = estim[(i-1)*12+3, 3]; x[i,4] = estim[(i-1)*12+4, 3];
  y[i,1] = estim[(i-1)*12+5, 3]; y[i,2] = estim[(i-1)*12+6, 3];
  y[i,3] = estim[(i-1)*12+7, 3]; y[i,4] = estim[(i-1)*12+8, 3];
  z[i,1] = estim[(i-1)*12+9, 3]; z[i,2] = estim[(i-1)*12+10, 3];
  z[i,3] = estim[(i-1)*12+11, 3]; z[i,4] = estim[(i-1)*12+12, 3];
END;

C = y; /* C : 범주별 실업자수 */
D = x + y; /* D : 범주별 경찰인구 */
B = D + z; /* B : 범주별 조사인구 */

x = A / B # C;
Ydir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) || SUM(x[,4]);
/* Ydir : 범주별 직접추정치 */
tA = SUM(A[,1]) || SUM(A[,2]) || SUM(A[,3]) || SUM(A[,4]);
/* A의 범주별 합계 */
DO i = 1 TO ngrps;

/* Ys : 합성추정값 */
  zYs = (A[i,] # Ydir) / tA;
  Ys[i] = SUM(zYs);
END;
USE direct1 VAR {jcode josagu jobtype edulevel count};
/* 경찰조사 데이터(조사구별) */
SETOUT outsyncom;
DO i = 1 TO ngrps;
/* 소지역 갯수 만큼 반복 */
  jcode2 = dirval[i,1];
  READ ALL WHERE (jcode = jcode2 & jobtype = 1) INTO
emp; /* 조사구별 취업자 데이터 */
  READ ALL WHERE (jcode = jcode2 & jobtype = 2) INTO
unemp; /* 조사구별 실업자 데이터 */
  READ ALL WHERE (jcode = jcode2 & jobtype = 3) INTO
noact; /* 조사구별 비경찰 데이터 */
  nofjosagu = NROW(emp) / 4;
  iYs = SHAPE(0, nofjosagu, 1);
  DO without = 1 TO nofjosagu;
/* 조사구 1개씩 제외하면서 B, C, D 재계산 */
    iB = B; iC = C; iD = D;
/* 제외하지 않은 B, C, D값 저장 */
    DO j = 1 TO 4;
/* 교육 범주 */
      dde[j] = emp[(without-1)*4+j, 5];
      ddu[j] = unemp[(without-1)*4+j, 5];
      ddn[j] = noact[(without-1)*4+j, 5];
    END;
  END;
END;

```

```

        END;
        iemp =dde'; iunemp =ddu'; inoact= ddn';
/* Transpose */
        B[i,] = B[i,] - (iemp + iunemp + inoact);
/* 조사구 1개 제외된 B, C, D */
        C[i,] = C[i,] - iunemp;
        x = A / B # C;
        iYdir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) ||
SUM(x[,4]);
/* iYdir : 조사구 제외후 범주별 직접추정치 */
        zYs = (A[i,] # iYdir) / tA;
        iYs[without] = SUM(zYs);
        B[i,] = B[i,] + (iemp + iunemp + inoact);
/* 제외 이전으로 B, C, D 복구 */
        C[i,] = C[i,] + iunemp;
END;

/* 합성추정값 계산 시작 */
avgYs = SUM(iYs) / nofjosagu;
varYs = iYs - avgYs;
varYs = varYs # varYs;
varYs = SUM(varYs) * (nofjosagu-1) / nofjosagu;
jcode = dirval[i,1];
synest = Ys[i];
synbias = (nofjosagu-1) * (avgYs - Ys[i]);
IF varYs < 0.0001 THEN DO;
/* 연산 오차 고려 */
        varYs = 0;
        synbias = 0;
END;
synvar = varYs;
mse = varYs + synbias * synbias;
synmse = SQRT(mse);
syncv = synmse / Ys[i] * 100;
/* 합성추정값 계산 완료 */

/* 복합추정값 계산 시작 */
direst = dirval[i,2];
dirvar = dirval[i,3];
dirvar = dirvar ** 2;
/* 직접추정치의 분산 */
IF (mse + dirvar) = 0 THEN wgt = 0;
ELSE wgt = mse / (mse + dirvar);
/* Weight 계산 */
Yc[i] = wgt * direst + (1-wgt) * Ys[i];
iYc = wgt * direst + (1-wgt) * iYs;
avgYc = SUM(iYc) / nofjosagu;
varYc = iYc - avgYc;

```

```

        varYc = varYc # varYc;
        varYc = SUM(varYc) * (nofjosagu-1) / nofjosagu;
        comest = Yc[i];
        combias = (nofjosagu - 1) * (avgYc - Yc[i]);
        IF varYc < 0.0001 THEN DO;
        /* 연산 오차 고려 */
            varYc = 0;
            combias = 0;
        END;
        comvar = varYc;
        commse = SQRT(varYc + combias * combias);
        comcv = commse / Yc[i] * 100;
        /* 복합추정값 계산 완료 */

        APPEND;
    END;

    SETOUT revised;
    DO i = 1 TO ngrps;
    /* 보정값 출력 */
        jcode = dirval[i,1];
        revsyn = SUM(dirval[,2]) * Ys[i] / SUM(Ys);
        revcom = SUM(dirval[,2]) * Yc[i] / SUM(Yc);
        APPEND;
    END;
    CLOSE outsyncom;
    CLOSE revised;
QUIT;

DATA &lib_2..outsyncom;
    SET outsyncom;
RUN;

DATA dir2;
    SET &lib_2..outdir;
    IF estunemp = 0 THEN dircv = 0;
    ELSE dircv = varunemp / estunemp * 100;
    KEEP jcode name estunemp varunemp dircv;
RUN;

PROC SQL;
    /* 계산 결과 테이블 조인 */
    CREATE TABLE tmpresult AS
    SELECT *
    FROM dir2 JOIN outsyncom ON dir2.jcode = outsyncom.jcode;
QUIT;

```

```

PROC SQL;
    CREATE TABLE result AS
    SELECT *
    FROM tmpresult JOIN revised ON tmpresult.jcode = revised.jcode;
QUIT;

DATA &outname;
    SET result;
    DROP jcode;
    LABEL name = '지역이름' estunemp = '실업자수[직접]'
           varunemp = '추정오차'
           dircv = 'CV[직접]'
           synest = '실업자수[합성]'
           synbias = 'Bias'
           synvar = '분 산'
           synmse = 'root.mse'
           syncv = 'CV[합성]'
           comest = '실업자수[복합]'
           combias = 'Bias'
           comvar = '분 산'
           commse = 'root.mse'
           comcv = 'CV[복합]'
           revsyn = '보 정 값[합성]'
           revcom = '보 정 값[복합]';

RUN;
%MEND;

%process(111, "d:\mySASpgm\0210\age\unemp\result\out1_11");
/* 15-29세 남자 (1집락: 특별시/광역시) */
%process(211, "d:\mySASpgm\0210\age\unemp\result\out2_11");
/* 15-29세 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(311, "d:\mySASpgm\0210\age\unemp\result\out3_11");
/* 15-29세 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(112, "d:\mySASpgm\0210\age\unemp\result\out1_12");
/* 15-29세 여자 (1집락: 특별시/광역시) */
%process(212, "d:\mySASpgm\0210\age\unemp\result\out2_12");
/* 15-29세 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(312, "d:\mySASpgm\0210\age\unemp\result\out3_12");
/* 15-29세 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(121, "d:\mySASpgm\0210\age\unemp\result\out1_21");
/* 30-39세 남자 (1집락: 특별시/광역시) */
%process(221, "d:\mySASpgm\0210\age\unemp\result\out2_21");
/* 30-39세 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(321, "d:\mySASpgm\0210\age\unemp\result\out3_21");
/* 30-39세 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(122, "d:\mySASpgm\0210\age\unemp\result\out1_22");

```

```

/* 30-39세 여자 (1집락: 특별시/광역시) */
%process(222, "d:\mySASpgm\0210\age\unemp\result\out2_22");
/* 30-39세 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(322, "d:\mySASpgm\0210\age\unemp\result\out3_22");
/* 30-39세 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(131, "d:\mySASpgm\0210\age\unemp\result\out1_31");
/* 40-49세 남자 (1집락: 특별시/광역시) */
%process(231, "d:\mySASpgm\0210\age\unemp\result\out2_31");
/* 40-49세 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(331, "d:\mySASpgm\0210\age\unemp\result\out3_31");
/* 40-49세 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(132, "d:\mySASpgm\0210\age\unemp\result\out1_32");
/* 40-49세 여자 (1집락: 특별시/광역시) */
%process(232, "d:\mySASpgm\0210\age\unemp\result\out2_32");
/* 40-49세 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(332, "d:\mySASpgm\0210\age\unemp\result\out3_32");
/* 40-49세 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(141, "d:\mySASpgm\0210\age\unemp\result\out1_41");
/* 50세 이상 남자 (1집락: 특별시/광역시) */
%process(241, "d:\mySASpgm\0210\age\unemp\result\out2_41");
/* 50세 이상 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(341, "d:\mySASpgm\0210\age\unemp\result\out3_41");
/* 50세 이상 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(142, "d:\mySASpgm\0210\age\unemp\result\out1_42");
/* 50세 이상 여자 (1집락: 특별시/광역시) */
%process(242, "d:\mySASpgm\0210\age\unemp\result\out2_42");
/* 50세 이상 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(342, "d:\mySASpgm\0210\age\unemp\result\out3_42");
/* 50세 이상 여자 (3집락: 강원, 충북, 전북, 전남) */

/* 처리 결과 통합 */

DATA &lib_3..out11;
    category = '남자 15-29세';
    SET &lib_2..out1_11 &lib_2..out2_11 &lib_2..out3_11;
    LABEL category = '연령범주';
RUN;

DATA &lib_3..out12;
    category = '여자 15-29세';
    SET &lib_2..out1_12 &lib_2..out2_12 &lib_2..out3_12;
    LABEL category = '연령범주';
RUN;

DATA &lib_3..out21;

```

```

        category = '남자 30-39세';
        SET &lib_2..out1_21 &lib_2..out2_21 &lib_2..out3_21;
        LABEL category = '연령범주';
RUN;

DATA &lib_3..out22;
        category = '여자 30-39세';
        SET &lib_2..out1_22 &lib_2..out2_22 &lib_2..out3_22;
        LABEL category = '연령범주';
RUN;

DATA &lib_3..out31;
        category = '남자 40-49세';
        SET &lib_2..out1_31 &lib_2..out2_31 &lib_2..out3_31;
        LABEL category = '연령범주';
RUN;

DATA &lib_3..out32;
        category = '여자 40-49세';
        SET &lib_2..out1_32 &lib_2..out2_32 &lib_2..out3_32;
        LABEL category = '연령범주';
RUN;

DATA &lib_3..out41;
        category = '남자 50이상';
        SET &lib_2..out1_41 &lib_2..out2_41 &lib_2..out3_41;
        LABEL category = '연령범주';
RUN;

DATA &lib_3..out42;
        category = '여자 50이상';
        SET &lib_2..out1_42 &lib_2..out2_42 &lib_2..out3_42;
        LABEL category = '연령범주';
RUN;

DATA total;
        SET &lib_3..out11 &lib_3..out12 &lib_3..out21 &lib_3..out22 &lib_3..out31
&lib_3..out32 &lib_3..out41 &lib_3..out42;
RUN;

%MACRO concat(city, cityname);
        DATA &city;
                SET total;
                IF name = &cityname;
RUN;

DATA cat;

```

```

        SET &city;
        KEEP category;
RUN;

PROC IML;
    CREATE t&city VAR {estunemp varunemp dircv synest synbias
synvar synmse syncv comest combias comvar commse comcv revsyn revcom};
    USE &city;
    READ ALL INTO indat;

    DO i = 1 TO NROW(indat);
        estunemp = indat[i, 1];
        varunemp = indat[i,2];
        dircv = indat[i, 3];
        synest = indat[i, 4];
        synbias = indat[i, 5];
        synvar = indat[i, 6];
        synmse = indat[i, 7];
        syncv = indat[i, 8];
        comest = indat[i, 9];
        combias = indat[i, 10];
        comvar = indat[i, 11];
        commse = indat[i, 12];
        comcv = indat[i, 13];
        revsyn = SUM(indat[, 1]) * comest / SUM(indat[, 4]);
        revcom = SUM(indat[, 1]) * comest / SUM(indat[, 9]);
    APPEND;

    END;
QUIT;

DATA t&city;
    SET cat;
    SET t&city;
    LABEL name = '지역이름' estunemp = '실업자수[직접]'
        varunemp = '추정오차'
        dircv = 'CV[직접]'
        synest = '실업자수[합성]'
        synbias = 'Bias'
        synvar = '분 산'
        synmse = 'root.mse'
        syncv = 'CV[합성]'
        comest = '실업자수[복합]'
        combias = 'Bias'
        comvar = '분 산'
        commse = 'root.mse'
        comcv = 'CV[복합]'
        revsyn = '보 정 값[합성]'

```



```

                                revcom = '보정값[복합]';
RUN;

PROC SORT data = t&city out = &lib_3.&city;
    BY category;
RUN;
%MEND;

/* 시도별로 묶은 후 보정값 계산 */

%concat(Seoul, "서울특별시");
%concat(Pusan, "부산광역시");
%concat(Taegu, "대구광역시");
%concat(Incheon, "인천광역시");
%concat(Gwangju, "광주광역시");
%concat(Daejeon, "대전광역시");
%concat(Ulsan, "울산광역시");
%concat(Kyonggi, "경기도");
%concat(Kangwon, "강원도");
%concat(Chbuk, "충청북도");
%concat(Chnam, "충청남도");
%concat(Jeonbuk, "전라북도");
%concat(Jeonnam, "전라남도");
%concat(Kyongbuk, "경상북도");
%concat(Kyongnam, "경상남도");
%concat(Jeju, "제주도");

/* 결과 출력 */

PROC PRINT DATA = &lib_3.Seoul LABEL;
    TITLE '서울특별시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Pusan LABEL;
    TITLE '부산광역시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Taegu LABEL;
    TITLE '대구광역시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Incheon LABEL;
    TITLE '인천광역시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Gwangju LABEL;
    TITLE '광주광역시 성별/연령별 실업자 현황';

```

```

RUN;

PROC PRINT DATA = &lib_3..Daejon LABEL;
    TITLE '대전광역시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Ulsan LABEL;
    TITLE '울산광역시 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Kyounggi LABEL;
    TITLE '경기도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Kangwon LABEL;
    TITLE '강원도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Chbuk LABEL;
    TITLE '충청북도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Chnam LABEL;
    TITLE '충청남도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Jeonbuk LABEL;
    TITLE '전라북도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Jeonnam LABEL;
    TITLE '전라남도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Kyongbuk LABEL;
    TITLE '경상북도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Kyongnam LABEL;
    TITLE '경상남도 성별/연령별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3..Jeju LABEL;
    TITLE '제주도 성별/연령별 실업자 현황';
RUN;

```

(2) 성별-교육정도별 추정 SAS 프로그램

```
/*
*****
경제활동인구 계산 프로그램(교육정도별/성별)
(새로운 조사구코드용, 실업자)
*****
*/

OPTIONS ls=132 ps=200 CENTER NODATE NONUMBER
FONT= 'Terminal' 8;

/*=====*/
%LET lib_1 = project;

/* dataset의 라이브러리명 */
%LET libpath_1 = 'd:\mySASpgm\0210\educat\unemp\sasdata';
/* 생성된dataset 저장 경로 */
%LET lib_2 = result ;

/* 생성된 dataset을 저장할 라이브러리명 */
%LET libpath_2 = 'd:\mySASpgm\0210\educat\unemp\result' ;
/* 생성된dataset의폴리직경로 */
%LET lib_3 = report;
%LET libpath_3 = 'd:\mySASpgm\0210\educat\unemp\report';
/*=====*/

LIBNAME &lib_1 &libpath_1 ;
LIBNAME &lib_2 &libpath_2 ;
LIBNAME &lib_3 &libpath_3 ;

/*=====
직접추정값 계산 루틴
=====*/

%MACRO process(category, outname);
DATA groupname; /* 시도이름 | 조사구 코드 */
SET &lib_1..popul&category;
KEEP name jcode;
RUN;

DATA inputdata; /* 데이터 파일 */
SET &lib_1..indat&category;
RUN ;

DATA direct1;
SET inputdata;
```

```

        popul = count * ratio;
RUN;

PROC MEANS
    /* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
    DATA = direct1 NOPRINT NWAY;
    OUTPUT OUT = direct2 SUM = popul;
    CLASS jcode jobtype;
    BY jcode;
    VAR popul;
RUN;

DATA direct2; /* 소지역번호 | 경제인 범주 | 직접추정치 */
    SET direct2;
    DROP josagu _TYPE_ _FREQ_;
RUN;

PROC IML SYMSIZE=1024;
    CREATE outdir VAR {jcode estemp varemp estunemp varunemp
estnoact varnoact};

    USE direct2 VAR {jcode jobtype popul}; /* 직접추정치 정보 */
    READ ALL INTO estim;
    USE direct1 VAR {josagu jobtype agetype count ratio jcode popul};
    /* 분산 계산위해 */

    nofgroups = NROW(estim) / 3;
    DO i = 1 TO nofgroups;
    /* 소지역 갯수 만큼 반복 */
        jcode2 = estim[i*3, 1];
        READ ALL WHERE (jcode = jcode2 & jobtype = 1) INTO
emp; /* 소지역의 취업자 데이터 */
        READ ALL WHERE (jcode = jcode2 & jobtype = 2) INTO
unemp; /* 소지역의 실업자 데이터 */
        READ ALL WHERE (jcode = jcode2 & jobtype = 3) INTO
noact; /* 소지역의 비경활 데이터 */

        nofjosagu = NROW(emp) / 4;
        /* 소지역의 조사구 갯수 */
        pivotrow = 0;
        IF emp[1,1]/100000 > 30 & MOD(emp[nofjosagu*4,1],
100000) >= 20000 THEN
            /* 시군 혼합 지역 (도) */
            DO j = nofjosagu*4 TO 4 BY -4;
                IF MOD(emp[j,1], 100000) >= 20000
THEN pivotrow = j;
                /* pivotrow : 읍면 조사구 시작 지점 */
            END;
        Zemp=SHAPE(0,nofjosagu,1);

```

```

/* 취업자 분산계산을 위해 */
Zunemp=SHAPE(0,nofjosagu,1);
/* 실업자 분산계산을 위해 */
Znoact = SHAPE(0, nofjosagu, 1);
/* 비경활의 분산 계산을 위해 */

totE = SUM(emp[,4]);
/* 취업자 합계 */
totU = SUM(unemp[,4]);
/* 실업자 합계 */
totN = SUM(noact[,4]);
/* 비경활인구 합계 */
totX = totE + totU;
/* 경활인구 합계 */
totT = totX + totN;
/* 조사인구 합계(경활+비경활) */

DO j = 1 TO nofjosagu - 1;
/* 조사구 수 만큼 반복 */
    josa1 = emp[(j-1)*4+1, 1];
/* (j) 번째 조사구 번호 */
    josa2 = emp[j*4+1, 1];
/* (j+1) 번째 조사구 번호 */
    READ ALL WHERE (josagu = josa1 & jobtype
= 1) INTO emp1;
/* j 조사구의 취업자 데이터 */
    READ ALL WHERE (josagu = josa1 & jobtype
= 2) INTO unemp1;
/* j 조사구의 실업자 데이터 */
    READ ALL WHERE (josagu = josa1 & jobtype
= 3) INTO noact1;
/* j 조사구의 비경활 데이터 */
    READ ALL WHERE (josagu = josa2 & jobtype
= 1) INTO emp2;
/* j+1 조사구의 취업자 데이터 */
    READ ALL WHERE (josagu = josa2 & jobtype
= 2) INTO unemp2;
/* j+1 조사구의 실업자 데이터 */
    READ ALL WHERE (josagu = josa2 & jobtype
= 3) INTO noact2;
/* j+1 조사구의 비경활 데이터 */

    Ej1 = SUM(emp1[,4]);
/* j 조사구의 취업자 합계 */
    Uj1 = SUM(unemp1[,4]);
/* j 조사구의 실업자 합계 */
    Nj1 = SUM(noact1[,4]);
/* j 조사구의 비경활인구 합계 */
    Xj1 = Ej1 + Uj1;
/* j 조사구의 경활인구 합계 */
    Tj1 = Xj1 + Nj1;
/* j 조사구의 조사인구 합계 */

```

```

        Ej2 = SUM(emp2[,4]);
/* j+1 조사구의 취업자 합계 */
        Uj2 = SUM(unemp2[,4]);
/* j+1 조사구의 실업자 합계 */
        Nj2 = SUM(noact2[,4]);
/* j+1 조사구의 비경활인구 합계 */
        Xj2 = Ej2 + Uj2;
/* j+1 조사구의 경활인구 합계 */
        Tj2 = Xj2 + Nj2;
/* j+1 조사구의 조사인구 합계 */

Zemp[j] = (Ej2 - Ej1) - ((Tj2 - Tj1) * totE / totT);
/* Zemp 계산 */
Zunemp[j] = (Uj2 - Uj1) - ((Tj2 - Tj1) * totU / totT);
/* Zunemp 계산 */
Znoact[j] = (Nj2 - Nj1) - ((Tj2 - Tj1) * totN / totT);
/* Znoact 계산 */
END;          /* 조사구 */

rat1 = emp[1,5];
/* 시지역 승수 */
IF pivotrow ^= 0 THEN DO;
        rat2 = emp[pivotrow,5];
/* 군지역 승수 */
        Zemp[pivotrow/4-1] = 0;
/* 접경지역의 Z 값 cut-off */
        Zunemp[pivotrow/4-1] = 0;
        Znoact[pivotrow/4-1] = 0;
END;

mtmpe = Zemp # Zemp;          /* 제공 */
mtmpu = Zunemp # Zunemp;     /* 제공 */
mtmpn = Znoact # Znoact;     /* 제공 */

IF pivotrow = 0 THEN nofjosagul = nofjosagu;
ELSE DO;
        nofjosagul = pivotrow / 4 - 1;
/* 시지역의 조사구 갯수 */
        nofjosagu2 = nofjosagu - pivotrow / 4 + 1;
/* 읍면의 조사구 갯수 */
END;

mul1 = rat1 * rat1 * nofjosagul / (2*(nofjosagul-1));
/* 승수와 조사구수 고려 */
tmpu = SUM(mtmpu);
/* 남자 취업자 Zemp의 제공 합계 */
tmpu = SUM(mtmpu);

```

```

/* 남자 실업자 */
tmpn = SUM(mtmpn);
/* 남자 비경활인 */
IF pivotrow = 0 THEN DO;
    tmpe = tmpe * null;
/* Zemp 값에 위의 값 곱하기 */
    tmpu = tmpu * null;
    tmpn = tmpn * null;
END;
ELSE DO;

/* 혼합조사구의 경우(읍면 조사구 합계) */
mul2 = rat2 * rat2 * nofjosagu2 / (2*(nofjosagu2-1));
/* 승수와 조사구수 고려 */
umpe = SUM(mtmpe[pivotrow/4-1:nofjosagu-1]);
/* 남자 취업자 Zemp의 제곱 합계 */
umpu = SUM(mtmpu[pivotrow/4-1:nofjosagu-1]);
/* 남자 실업자 */
umpn = SUM(mtmpn[pivotrow/4-1:nofjosagu-1]);
/* 남자 비경활인 */
tmpe = (tmpe-umpe) * null + umpe * mul2;
tmpu = (tmpu-umpu) * null + umpu * mul2;
tmpn = (tmpn-umpn) * null + umpn * mul2;
END;

jcode = jcode2;
varemp = SQRT(tmpe);
varunemp = SQRT(tmpu);
varnoact = SQRT(tmpn);
estemp = estim[(i-1)*3+1, 3];
estunemp = estim[(i-1)*3+2, 3];
estnoact = estim[(i-1)*3+3, 3];
APPEND;
/* 취업자 추정값 및 분산, 실업자 추정값 및 분산, 비경활인 추정값 및 분산 출력 */
END;
/* 소지역 */
CLOSE outdir;
QUIT;

PROC SQL;
/* 직접추정값 출력 */
CREATE TABLE &lib_2.outdir AS
/* 시군이름 | 조사구 코드 | 취업자 추정값 | 분산 | 실업자 추정값 | 분산 | 비경활 추정값 | 분산 */
SELECT *
FROM groupname JOIN outdir ON groupname.jcode = outdir.jcode;
QUIT;

```

```

/*=====
합성 및 복합추정값(제크나이프) 계산 루틴
=====*/

```

```
PROC MEANS
```

```

/* 소지역별 직접추정치(취업자수, 실업자수, 비경활인수) 계산 */
DATA = direct1 NOPRINT NWAY;
OUTPUT OUT = direct2 SUM = count popul;
CLASS jcode jobtype agetype;
BY jcode;
VAR count popul;

```

```
RUN;
```

```

DATA direct2; /* 소지역번호 | 경제인 범주 | 직접추정치 */
SET direct2;
DROP josagu _TYPE_ _FREQ_;

```

```
RUN;
```

```
PROC IML;
```

```

CREATE outsyncom VAR {jcode synest synbias synvar synmse
syncv comest combias comvar commse comcv};
CREATE revised VAR {jcode revsyn revcom};

```

```

ctgry4);
USE &lib_1..popul&category VAR {jcode ctgry1 ctgry2 ctgry3
/* 상주추정인구 파일 */
READ ALL VAR {ctgry1 ctgry2 ctgry3 ctgry4} INTO A;
/* A : 상주추정인구 */
USE direct2 VAR {jcode jobtype count};

```

```

/* 경찰조사 데이터(소지역별) */
READ ALL INTO estim;
USE &lib_2..outdir VAR {jcode estunemp varunemp};
/* 직접추정치 및 분산 입력 */
READ ALL INTO dirval;

```

```

ngrps = NROW(dirval); /* 시도 갯수 */
x = SHAPE(0, ngrps, 4); /* 범주별 취업자수 */
y = SHAPE(0, ngrps, 4); /* 범주별 실업자수 */
z = SHAPE(0, ngrps, 4); /* 범주별 비경활인수 */
dde = SHAPE(0, 4, 1);
ddu = SHAPE(0, 4, 1);
ddn = SHAPE(0, 4, 1);
Ys = SHAPE(0, ngrps, 1);
Yc = SHAPE(0, ngrps, 1);

```



```

DO i = 1 TO ngrps;
    x[i,1]=estim[(i-1)*12+1, 3]; x[i,2]=estim[(i-1)*12+2, 3];
    x[i,3]=estim[(i-1)*12+3, 3]; x[i,4]=estim[(i-1)*12+4, 3];
    y[i,1]=estim[(i-1)*12+5, 3]; y[i,2]=estim[(i-1)*12+6, 3];
    y[i,3]=estim[(i-1)*12+7, 3]; y[i,4]=estim[(i-1)*12+8, 3];
    z[i,1]=estim[(i-1)*12+9, 3]; z[i,2]=estim[(i-1)*12+10, 3];
    z[i,3]=estim[(i-1)*12+11, 3]; z[i,4]=estim[(i-1)*12+12, 3];
END;

C = y;          /* C : 범주별 실업자수 */
D = x + y;     /* D : 범주별 경찰인구 */
B = D + z;     /* B : 범주별 조사인구 */

x = A / B # C;
Ydir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) || SUM(x[,4]);
/* Ydir : 범주별 직접추정치 */
tA = SUM(A[,1]) || SUM(A[,2]) || SUM(A[,3]) || SUM(A[,4]);
/* A의 범주별 합계 */
DO i = 1 TO ngrps;

/* Ys : 합성추정값 */
    zYs = (A[i,] # Ydir) / tA;
    Ys[i] = SUM(zYs);
END;
USE direct1 VAR (jcode josagu jobtype agetype count);
/* 경찰조사 데이터(조사구별) */
SETOUT outsyncom;
    DO i = 1 TO ngrps;
/* 소지역 갯수 만큼 반복 */
        jcode2 = dirval[i,1];
        READ ALL WHERE (jcode = jcode2 & jobtype = 1) INTO
emp; /* 조사구별 취업자 데이터 */
        READ ALL WHERE (jcode = jcode2 & jobtype = 2) INTO
unemp; /* 조사구별 실업자 데이터 */
        READ ALL WHERE (jcode = jcode2 & jobtype = 3) INTO
noact; /* 조사구별 비경찰 데이터 */
        nofjosagu = NROW(emp) / 4;
        iYs = SHAPE(0, nofjosagu, 1);
        DO without = 1 TO nofjosagu;
/* 조사구 1개씩 제외하면서 B, C, D 재계산 */
            iB = B; iC = C; iD = D;
/* 제외하지 않은 B, C, D값 저장 */
            DO j = 1 TO 4;
                dde[j] = emp[(without-1)*4+j, 5];
                ddu[j] = unemp[(without-1)*4+j, 5];
                ddn[j] = noact[(without-1)*4+j, 5];
            END;
        END;
    END;

```

```

END;
iemp=dde'; iunemp=ddu'; inoact=ddn';
/* Transpose */
B[i,] = B[i,] - (iemp + iunemp + inoact);
/* 조사구 1개 제외된 B, C, D */
C[i,] = C[i,] - iunemp;
x = A / B # C;
SUM(x[,4]);
iYdir = SUM(x[,1]) || SUM(x[,2]) || SUM(x[,3]) ||
/* iYdir : 조사구 제외후 범주별 직접추정치 */
zYs = (A[i,] # iYdir) / tA;
iYs[without] = SUM(zYs);
B[i,] = B[i,] + (iemp + iunemp + inoact);
/* 제외 이전으로 B, C, D 복구 */
C[i,] = C[i,] + iunemp;
END;

/* 합성추정값 계산 시작 */
avgYs = SUM(iYs) / nofjosagu;
varYs = iYs - avgYs;
varYs = varYs # varYs;
varYs = SUM(varYs) * (nofjosagu-1) / nofjosagu;
jcode = dirval[i,1];
synest = Ys[i];
synbias = (nofjosagu-1) * (avgYs - Ys[i]);
IF varYs < 0.0001 THEN DO;
/* 연산 오차 고려 */
varYs = 0;
synbias = 0;
END;
synvar = varYs;
mse = varYs + synbias * synbias;
synmse = SQRT(mse);
syncv = synmse / Ys[i] * 100;
/* 합성추정값 계산 완료 */

/* 복합추정값 계산 시작 */
direst = dirval[i,2];
dirvar = dirval[i,3];
dirvar = dirvar ** 2;
/* 직접추정치의 분산 */
IF (mse + dirvar) = 0 THEN wgt = 0;
ELSE wgt = mse / (mse + dirvar);
/* Weight 계산 */
Yc[i] = wgt * direst + (1-wgt) * Ys[i];
iYc = wgt * direst + (1-wgt) * iYs;
avgYc = SUM(iYc) / nofjosagu;
varYc = iYc - avgYc;

```

```

varYc = varYc # varYc;
varYc = SUM(varYc) * (nofjosagu-1) / nofjosagu;
comest = Yc[i];
combias = (nofjosagu - 1) * (avgYc - Yc[i]);
IF varYc < 0.0001 THEN DO;
/* 연산 오차 고려 */
varYc = 0;
combias = 0;
END;
comvar = varYc;
commse = SQRT(varYc + combias * combias);
comcv = commse / Yc[i] * 100;
/* 복합추정값 계산 완료 */

APPEND;
END;

SETOUT revised;
DO i = 1 TO ngrps;
/* 보정값 출력 */
jcode = dirval[i,1];
revsyn = SUM(dirval[,2]) * Ys[i] / SUM(Ys);
revcom = SUM(dirval[,2]) * Yc[i] / SUM(Yc);
APPEND;
END;
CLOSE outsyncom;
CLOSE revised;
QUIT;

DATA &lib_2.outsyncom;
SET outsyncom;
RUN;

DATA dir2;
SET &lib_2.outdir;
IF estunemp = 0 THEN dircv = 0;
ELSE dircv = varunemp / estunemp * 100;
KEEP jcode name estunemp varunemp dircv;
RUN;

PROC SQL;
/* 계산 결과 테이블 조인 */
CREATE TABLE tmpresult AS
SELECT *
FROM dir2 JOIN outsyncom ON dir2.jcode = outsyncom.jcode;
QUIT;

```

```

PROC SQL;
    CREATE TABLE result AS
    SELECT *
    FROM tmpresult JOIN revised ON tmpresult.jcode = revised.jcode;
QUIT;

DATA &outname;
    SET result;
    DROP jcode;
    LABEL name = '지역이름' estunemp = '실업자수[직접]'
           varunemp = '추정오차'   dircv = 'CV[직접]'
           synest = '실업자수[합성]' synbias = 'Bias'
           synvar = '분산'         synmse = 'root.mse'
           syncv = 'CV[합성]'     comest = '실업자수[복합]'
           combias = 'Bias'       comvar = '분산'
           commse = 'root.mse'    comcv = 'CV[복합]'
           revsyn = '보정값[합성]' revcom = '보정값[복합]';

RUN;
%MEND;

%process(111, "d:\mySASpgm\0210\educat\unemp\result\out1_11");
/* 초등학교 졸업 이하 남자 (1집락: 특별시/광역시) */
%process(211, "d:\mySASpgm\0210\educat\unemp\result\out2_11");
/* 초등학교 졸업 이하 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(311, "d:\mySASpgm\0210\educat\unemp\result\out3_11");
/* 초등학교 졸업 이하 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(112, "d:\mySASpgm\0210\educat\unemp\result\out1_12");
/* 초등학교 졸업 이하 여자 (1집락: 특별시/광역시) */
%process(212, "d:\mySASpgm\0210\educat\unemp\result\out2_12");
/* 초등학교 졸업 이하 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(312, "d:\mySASpgm\0210\educat\unemp\result\out3_12");
/* 초등학교 졸업 이하 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(121, "d:\mySASpgm\0210\educat\unemp\result\out1_21");
/* 중학교 졸업 남자 (1집락: 특별시/광역시) */
%process(221, "d:\mySASpgm\0210\educat\unemp\result\out2_21");
/* 중학교 졸업 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(321, "d:\mySASpgm\0210\educat\unemp\result\out3_21");
/* 중학교 졸업 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(122, "d:\mySASpgm\0210\educat\unemp\result\out1_22");
/* 중학교 졸업 여자 (1집락: 특별시/광역시) */
%process(222, "d:\mySASpgm\0210\educat\unemp\result\out2_22");
/* 중학교 졸업 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(322, "d:\mySASpgm\0210\educat\unemp\result\out3_22");
/* 중학교 졸업 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(131, "d:\mySASpgm\0210\educat\unemp\result\out1_31");

```

```

/* 고등학교 졸업 남자 (1집락: 특별시/광역시) */
%process(231, "d:\mySASpgm\0210\educat\unemp\result\out2_31");
/* 고등학교 졸업 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(331, "d:\mySASpgm\0210\educat\unemp\result\out3_31");
/* 고등학교 졸업 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(132, "d:\mySASpgm\0210\educat\unemp\result\out1_32");
/* 고등학교 졸업 여자 (1집락: 특별시/광역시) */
%process(232, "d:\mySASpgm\0210\educat\unemp\result\out2_32");
/* 고등학교 졸업 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(332, "d:\mySASpgm\0210\educat\unemp\result\out3_32");
/* 고등학교 졸업 여자 (3집락: 강원, 충북, 전북, 전남) */

%process(141, "d:\mySASpgm\0210\educat\unemp\result\out1_41");
/* 대학교 졸업 이상 남자 (1집락: 특별시/광역시) */
%process(241, "d:\mySASpgm\0210\educat\unemp\result\out2_41");
/* 대학교 졸업 이상 남자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(341, "d:\mySASpgm\0210\educat\unemp\result\out3_41");
/* 대학교 졸업 이상 남자 (3집락: 강원, 충북, 전북, 전남) */
%process(142, "d:\mySASpgm\0210\educat\unemp\result\out1_42");
/* 대학교 졸업 이상 여자 (1집락: 특별시/광역시) */
%process(242, "d:\mySASpgm\0210\educat\unemp\result\out2_42");
/* 대학교 졸업 이상 여자 (2집락: 경기, 충남, 경북, 경남, 제주) */
%process(342, "d:\mySASpgm\0210\educat\unemp\result\out3_42");
/* 대학교 졸업 이상 여자 (3집락: 강원, 충북, 전북, 전남) */

/* 처리 결과 통합 */

DATA &lib_3.out11;
    category = '1. 남자 초졸이하';
    SET &lib_2.out1_11 &lib_2.out2_11 &lib_2.out3_11;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out12;
    category = '5. 여자 초졸이하';
    SET &lib_2.out1_12 &lib_2.out2_12 &lib_2.out3_12;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out21;
    category = '2. 남자 중졸';
    SET &lib_2.out1_21 &lib_2.out2_21 &lib_2.out3_21;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out22;
    category = '6. 여자 중졸';

```

```

        SET &lib_2.out1_22 &lib_2.out2_22 &lib_2.out3_22;
        LABEL category = '교육수준';
RUN;

DATA &lib_3.out31;
    category = '3. 남자 고졸';
    SET &lib_2.out1_31 &lib_2.out2_31 &lib_2.out3_31;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out32;
    category = '7. 여자 고졸';
    SET &lib_2.out1_32 &lib_2.out2_32 &lib_2.out3_32;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out41;
    category = '4. 남자 대졸이상';
    SET &lib_2.out1_41 &lib_2.out2_41 &lib_2.out3_41;
    LABEL category = '교육수준';
RUN;

DATA &lib_3.out42;
    category = '8. 여자 대졸이상';
    SET &lib_2.out1_42 &lib_2.out2_42 &lib_2.out3_42;
    LABEL category = '교육수준';
RUN;

DATA total;
    SET &lib_3.out11 &lib_3.out12 &lib_3.out21 &lib_3.out22 &lib_3.out31
    &lib_3.out32 &lib_3.out41 &lib_3.out42;
RUN;

%MACRO concat(city, cityname);
    DATA &city;
        SET total;
        IF name = &cityname;
    RUN;

    DATA cat;
        SET &city;
        KEEP category;
    RUN;

    PROC IML;
        CREATE t&city VAR (estunemp varunemp dircv synest synbias
synvar synmse syncv comest combias comvar commse comcv revsyn revcom);

```

```

USE &city;
READ ALL INTO indat;

DO i = 1 TO NROW(indat);
    estunemp = indat[i, 1];
    varunemp = indat[i,2];
    dircv = indat[i, 3];
    synest = indat[i, 4];
    synbias = indat[i, 5];
    synvar = indat[i, 6];
    synmse = indat[i, 7];
    syncv = indat[i, 8];
    comest = indat[i, 9];
    combias = indat[i, 10];
    comvar = indat[i, 11];
    commse = indat[i, 12];
    comcv = indat[i, 13];
    revsyn = SUM(indat[, 1]) * comest / SUM(indat[, 4]);
    revcom = SUM(indat[, 1]) * comest / SUM(indat[, 9]);
    APPEND;
END;
QUIT;

DATA t&city;
    SET cat;
    SET t&city;
    LABEL name = '지역이름'
           estunemp = '실업자수[직접]'   varunemp = '추정오차'
           dircv = 'CV[직접]'   synest = '실업자수[합성]'
           synbias = 'Bias'     synvar = '분 산'
           synmse = 'root.mse'   syncv = 'CV[합성]'
           comest = '실업자수[복합]'   combias = 'Bias'
           comvar = '분 산'     commse = 'root.mse'
           comcv = 'CV[복합]'   revsyn = '보 정 값[합성]'
           revcom = '보 정 값[복합]';

RUN;

PROC SORT data = t&city out = &lib_3..&city;
    BY category;
RUN;
%MEND;

/* 시도별로 묶은 후 보정값 계산 */

%concat(Seoul, "서울특별시");
%concat(Pusan, "부산광역시");
%concat(Taegu, "대구광역시");

```

```

%concat(Incheon, "인천광역시");
%concat(Gwangju, "광주광역시");
%concat(Daejon, "대전광역시");
%concat(Ulsan, "울산광역시");
%concat(Kyounggi, "경기도");
%concat(Kangwon, "강원도");
%concat(Chbuk, "충청북도");
%concat(Chnam, "충청남도");
%concat(Jeonbuk, "전라북도");
%concat(Jeonnam, "전라남도");
%concat(Kyongbuk, "경상북도");
%concat(Kyongnam, "경상남도");
%concat(Jeju, "제주도");

```

```

/* 결과 출력 */

```

```

PROC PRINT DATA = &lib_3.Seoul LABEL;
    TITLE '서울특별시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Pusan LABEL;
    TITLE '부산광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Taegu LABEL;
    TITLE '대구광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Incheon LABEL;
    TITLE '인천광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Gwangju LABEL;
    TITLE '광주광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Daejon LABEL;
    TITLE '대전광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Ulsan LABEL;
    TITLE '울산광역시 성별/교육정도별 실업자 현황';
RUN;

```

```

PROC PRINT DATA = &lib_3.Kyounggi LABEL;
    TITLE '경기도 성별/교육정도별 실업자 현황';
RUN;

```



```

PROC PRINT DATA = &lib_3.Kangwon LABEL;
    TITLE '강원도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Chbuk LABEL;
    TITLE '충청북도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Chnam LABEL;
    TITLE '충청남도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Jeonbuk LABEL;
    TITLE '전라북도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Jeonnam LABEL;
    TITLE '전라남도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Kyongbuk LABEL;
    TITLE '경상북도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Kyongnam LABEL;
    TITLE '경상남도 성별/교육정도별 실업자 현황';
RUN;

PROC PRINT DATA = &lib_3.Jeju LABEL;
    TITLE '제주도 성별/교육정도별 실업자 현황';
RUN;

```