

2003년 노동부의 「소지역 추정기법을 이용한 임금통계 개선방안」에 관한 학술연구용역사업에 의한 것임

소지역추정기법에 의한 임금통계조사 개선 방안 최종보고서

2003. 12.

한국조사연구학회

<요약문>

(1) 현행 임금구조통계조사 자료 분석

▶ 산업소분류별 표본현황 및 소지역 추정기법의 필요성

- ▷ 산업소분류별 평균임금에 있어서 상당한 차이를 보임
- ▷ 일부 산업소분류의 경우 조사완료 표본 근로자수 또는 사업체수가 매우 작음
 - 통계의 정확도 확보를 위해서는 표본 근로자수 보다 사업체수가 중요함
- ▷ 표본 사업체가 전혀 없는 산업소분류도 있음
 - 소지역 추정대상에 포함여부 검토 필요함

▶ 직종분류별 표본 현황 및 소지역 추정법의 필요성

- ▷ 연구대상인 직종세분류별 표본현황 및 임금특성은 파악 불가능
 - 2003년 조사에서 실험적인 직종세분류 조사 필요함
- ▷ 직종소분류별 평균임금에 있어서 상당한 차이를 보임
- ▷ 소분류별 표본현황을 참고로 하면 상당수 직종세분류별 표본 확보에 어려움 예상됨
 - 통계의 정확도 확보를 위해서는 표본 근로자수 보다 사업체수가 중요함
 - 표본 사업체가 전혀 없는 산업세분류: 소지역 추정대상에 포함여부 검토 필요함

▶ 산업분류내 직종중(소)분류에 따른 표본 현황 및 평균임금

- ▷ 동일 산업중분류내에서도 근로자의 직종에 따른 임금에 격차 큼
 - 산업소분류별 소지역 추정모형에서 직종분류의 활용 필요함
- ▷ 산업소분류내 직종분류별 구성비에 대한 정확한 보조정보 확보 가능성 검토

(2) 산업소분류 평균임금 산출을 위한 소지역 추정기법

▶ 소지역추정 모형 구현 과정

- 관심변수인 근로자의 평균임금을 효율적으로 설명할 수 있는 보조변수에 대한 분석
- 다른 통계 또는 전수조사에서 선택된 보조변수와 관련된 정보의 확보 가능성 검토
- 선택된 보조변수를 효과적으로 활용할 수 있는 간접추정방법의 구현

○ 직접추정량과 간접추정량을 활용한 효율적인 복합추정량(composite estimator)의 구현

※ 기초분석 대상 주요 산업중분류

중분류 15 : 음식료품 제조업

중분류 17 : 섬유제품 제조업

중분류 24 : 화합물 및 화학제품 제조업

중분류 32 : 영상, 음향 및 통신장비 제조업

▶ 평균임금 산출 모형구현을 위한 보조변수 선택

▷ 상관분석을 통해 관련 변수 탐색

▷ 의사결정나무 분석의 활용

▷ 회귀분석의 변수선택과성을 통한 분석

=> 최종 선택된 분석대상 설명변수 참조 (임금 : $\log(\text{임금총액})$ 사용)

▶ 다른 통계조사의 보조정보 활용 방안 검토

▷ 검토대상 : 사업체노동실태조사, 노동력수요동향조사, 사업체기초통계 등

▷ 예상 모형관련 설명변수를 기준으로 노동력수요동향조사 활용도 점검 필요

▷ 노동력수용동향조사 활용을 위한 모형 적합도 검토 결과

- 활용 가능 변수: 근로자수, 성별구성비율, 직종분류

- 활용 가능 변수만을 활용한 모형은 적합도가 너무 낮다는 분석결과 도출

- 노동력수요동향조사의 보조정보를 모형에 부분적으로 반영하는 방안 검토

▶ 소지역 추정을 위한 선형모형의 적합

▷ 산업소분류 군집(중분류 또는 평균임금)별 최적 선형모형 검토

▷ 소분류별 적합에 따른 변동성 보안 가능성 검토 (로버스트한 모형 설정 목적)

▷ 일부 범주형 변수 중간값 사용 여부, 직종분류 반영시 분류 수준(중/소) 검토

▶ 간접추정량 산출 방안

▷ 산업중분류별 선형모형 활용 (범주사용, 직종중분류 사용)

▷ 산업중분류별 회귀모형을 기초로 산업소분류별 간접추정값 산출

- 소분류별 근로자 근속년수, 직종분류, 성별, 사업체규모 등의 특성 반영

- 평균(연속형), 또는 구성비(범주형) 활용 방법 검토

- 노동력수용동향 조사 등의 보조정보 부분적 활용 방안 검토

- 제시된 간접추정 모형 관련 구체적인 이론 유도 및 검증 (기존 방법과 다름)

▶ **복합추정량의 도출**

- ▷ 직접추정량과 간접추정량의 가중평균 형태의 복합추정량 사용
- ▷ 복합추정량의 가중값 결정
 - 이론적인 최적 가중값 보다 표본크기의존 가중값이 보다 현실적
- ▷ 복합추정량의 오차 추정 문제
 - 잭나이프 추정법 또는 분산함수 활용
- ▷ 향후 충분한 이론 개발 및 효율성 검증 작업이 필요함

▶ **직중세분류에 대한 소지역 추정기법 도입 방안**

- ▷ 고용구조조사 결과 사용
- ▷ 일반화선형모형 이용
- ▷ 향후 충분한 이론 개발 및 효율성 검증 작업이 필요함

(3) 향후 연구 방향

- ▷ 외부 보조정보의 활용가능성에 대한 심층적인 분석
- ▷ 선형모형에서 얻어진 간접추정방법의 효율성 검토 및 보완
- ▷ 복합추정량의 표본오차 추정 방안 검토
- ▷ 직중세분류에 대한 소지역 추정기법 도입 방안 검토

< 차례 >

제 1 장 연구개요

- 1.1 연구목적 및 필요성
- 1.2 연구내용 및 방법
- 1.3 기대효과
- 1.4 연구협의회 토의 사항 정리

제 2 장 소지역 추정법(Small Area Estimation)

- 2.1 직접 추정법(Direct Estimation)
- 2.2 간접 추정법(Indirect Estimation)
 - 2.2.1 인구통계적 방법(Demographic Method)
 - 2.2.2 합성 추정법(Synthetic Estimation)
 - 2.2.3 복합 추정법(Composite Estimation)
- 2.3 모형 기반 추정법(Model-Based Estimation)
 - 2.3.1 기본적인 지역 수준 모형(Basic Area-level Model)
 - 2.3.2 경험적 최량선형불편예측(EBLUP) 방법
 - 2.3.3 경험적 베이지스(EB) 방법
 - 2.3.4 계층적 베이지스(HB) 방법
- 2.4 소지역 평균의 추정량
- 2.5 외국의 소지역 추정법 적용 사례
 - 2.5.1 미국 주단위와 지역단위의 취업통계
 - 2.5.2 OES데이터에 대한 소지역 추정법 적용사례

제 3 장 현행 임금구조기통계조사 표본 분석

- 3.1 현행 임금구조통계조사 표본설계 개요
- 3.2 현행 임금구조통계조사 자료 분석
- 3.3 현행 임금구조통계조사에 대한 평가

제 4 장 산업소분류 평균임금 산출을 위한 소지역 추정기법

- 4.1 상관분석
- 4.2 선형모형 구현을 위한 설명변수 선택
- 4.3 다른 통계조사의 보조정보 활용 방안 검토
- 4.4 소지역 추정을 위한 선형모형의 적합
- 4.5 산업소분류 평균임금 추정을 위한 복합추정
- 4.6 산업소분류 평균임금 추정을 위한 EBLUP 추정
- 4.7 산업소분류 평균임금 추정결과 비교 및 향후 과제

제 5 장 직종세분류 평균임금 산출을 위한 소지역 추정기법

- 5.1 고용구조통계조사 개요
- 5.2 고용구조조사 결과를 이용한 직종세분류별 평균임금 추정 연구방향
- 5.3 소지역 추정을 위한 선형모형의 적합
- 5.4 직종세분류별 평균임금 추정

제 6 장 향후 연구방향(2004년 연구계획)

참고문헌

제 1 장 연구개요

1.1 연구목적 및 필요성

- 인력수급전망 인프라 구축을 위해 현행 임금구조기본통계조사의 산업중분류, 직종중(소)분류 단위 분석체계를 산업소분류, 직종세분류 수준까지 확대 생산하는 방안을 모색하기 위함

⇒ '06년부터 상기 분류기준에 따른 임금구조기본통계조사 결과 제공 목표

- 산업소분류, 직종세분류별 추정 근로자수, 임금 등

1.2 연구내용 및 방법

(1) 연구내용

① 소지역 통계작성을 위한 임금구조통계의 정비방안

- 통계조사항목 보완에 대한 제안
- 정도 높은 추정을 위한 기초통계의 안정적 생산 방안
 - . 추정정보로서 현행표본의 대표성 및 표본규모의 적합 여부 검토
 - . 통계조사 방법(현행 자계식)의 적정성 검토
 - . 직업분류(현행 ILO의 ISCO분류체계) 적용의 개선방안

② 세부영역 단위의 근로자수와 임금의 추정법

- 산업소분류 구분에 따라 근로자 수, 근로시간과 임금 등을 추정할 수 있는 소지역 추정법 연구
- 직종세분류 구분에 따라 근로자 수, 근로시간과 임금 등을 추정할 수 있는 소지역 추정기법 연구
- 근로자 수가 적은 희소산업 또는 희소 직종에 대한 안정적인 추정방안 연구
- 기존 조사자료(고용보험 통계, 노동력수요동향조사, 사업체기초통계조사 등)를 보조정보로 이용하여 세분된 분석단위 통계의 신뢰성을 높이는 방안

③ 추정의 정도를 높이기 위한 보조자료 활용 계획

- 노동력수요동향(현행 산업·직종구분류, 2004년 이후 신분류로 개편), 고용구조 통계(중앙고용정보원, WIC-OES 직업분류체계 이용) 등의 활용가능성

④ 통계인프라 구축 방안

- 통계조사담당 직원에 대한 조사의 전문성 확보 방안
- 조사내용의 질적인 향상을 위한 소요 예산 및 행정지원의 수준

<참고> 노동부의 향후 추진 계획

- 2004 ~ 2005년도(3단계)
 - 소지역단위 통계조사 실시
.산업세분류(3-digit) 및 직종세분류(4-digit)
 - '03 ~ '04년도 조사결과를 이용한 추정프로그램의 정도(Precision)검증 및 보완
- 2006년도 이후 (4단계)
 - 2005년 이후의 소지역 추정통계 생산으로 산업소분류 및 직종세분류 단위의 인력수급전망에 활용

(2) 연구방법

(가) 현행 임금구조기본통계조사 결과에 대한 실태분석

- 세분된 분석단위의 통계생산에 알맞도록 현행 표본설계와 조사방법에 대한 문제점 분석과 보완대책 연구
- 조사항목의 적절성 검토와 자료수집법의 타당성 분석
- 고용보험통계와 사업체기초통계조사 등의 보조정보를 이용할 수 있도록 전체적인 통계작성 절차에 대한 검토

(나) 세분된 분석단위에 대한 노동통계 생산을 위한 소지역 추정법에 이룬 연구

- 세분된 분석단위에 대한 근로자 수, 근로시간과 임금 등의 노동통계 생산을 위한 소지역 추정법의 이론과 알고리즘 연구
- 임금관련 통계조사에 대한 선진국(미국, 캐나다 등)의 통계생산 체계와 소지역 추정법 적용에 대한 사례분석
- 국내외의 소지역 추정기법을 적용한 사례연구 분석(노동력이나 임금통계 관련 사례)

(다) 산업소분류별 통계를 생산할 수 있는 소지역 추정법의 적용 연구

- 산업분류별(산업대분류 또는 산업중분류) 근로자수와 임금통계 구조분석
.사업체기초통계조사와 임금구조통계조사 자료를 연계하여 사업체 규모별 특성파악
.산업소분류의 군집화 방안 연구
- 소지역 추정기법 연구
.각 산업중분류에서 소분류와 근로자 수에 따라 군집화 방안 연구
.임금구조기본통계조사에서 구한 산업소분류 단위의 직접추정값(direct estimates)과 사업체기초통계조사와 임금구조통계조사에서 구한 합성추정값(synthetic estimates)을 결합한 복합추정값(composite estimates) 계산을 위한 알고리즘 개발
.복합추정값에서 직접추정값과 합성추정값에 부여할 최적 가중값의 산정 방안 연구
- 대표적인 산업대분류(제조업 또는 도.소매업 등)에 대하여 산업소분류별 근로자 수와 임금통계 작성을 위한 알고리즘 개발과 타당성 검토

- . 소지역 노동관련 통계에 대한 복합추정값과 추정오차를 계산할 수 있는 알고리즘 개발
- . 최근의 노동통계조사 결과를 대상으로 개발된 알고리즘을 검증하기 위한 사례분석
- . 타 산업분류로 일반화하는 방안 연구

(라) 직종세분류별 통계생산 기법개발과 알고리즘 연구

- 고용보험통계와 임금구조통계조사 자료를 이용한 임금, 근로자 수 등에 대한 구조분석
- 직종소분류별로 사업체 규모와 근로자수를 이용하여 합성추정방법을 적용할 수 있도록 유사한 집단의 군집화 방안 연구
- 직종소분류에 따라 적용할 수 있는 적절한 소지역 추정법의 모형화와 알고리즘 개발
- 중앙고용정보원과 통계청의 자료를 이용하여 직종별로 소지역 추정법으로 각종 노동관련 통계 생산을 위한 프로그램개발
- 소지역 추정기법 알고리즘 구현을 프로그램 작성과 실제자료를 이용한 타당성 검토

(마) 임금구조기본통계조사의 세분된 분석단위별 통계를 체계적으로 생산할 수 있는

종합적인 프로그램 개발 연구 및 정책적인 제안

- 연구 개발된 산업소분류 및 직종세분류 단위의 노동통계 추정법의 효율성과 신뢰성을 평가하기 위한 평균제곱오차(MSE : Mean Squared Error)의 계산 방법을 연구하고, 이러한 통계를 체계적으로 생산할 수 있는 정책적 제안을 한다.

1.3 기대효과

(1) 현행 임금구조기본통계조사의 분석체계를 확대 개선하여 다양한 분석단위(산업소분류, 직종세분류 등)의 노동통계를 생산할 수 있는 기반 마련

==> 인력수급전망 인프라 구축에 필요한 산업별, 직종별 인력수급 전망에 대한 예측값 산출

(2) 산업소분류와 직종세분류 단위의 노동통계를 소지역 추정법을 이용하여 생산함으로써 경제성과 신뢰성을 높일 수 있으며, 노동정책 수립과 집행의 기초 정보를 제공할 뿐 아니라 시행 결과의 평가 자료로 활용 가능할 것이다.

(3) 산업소분류와 직종세분류 단위의 정확한 노동통계를 생산함으로써 전반적으로 노동통계의 활용성과 실용성을 제고할 수 있고, 사업체기초통계조사와 고용보험 데이터베이스를 적극 활용함으로써 관리체계를 정립하여 노동통계 관련 정보체계의 효율성과 실용성을 제고할 수 있을 것이다.

(4) 세분화된 직종별, 산업별 소요 전문인력을 장기적으로 예측하여 맞춤형 교육과정 등을 운영함으로써 교육의 경쟁력을 높일 수 있을 것이다.

1.4 연구협의회 토의 사항 정리

(1) 1차 중간연구협의회 토의 사항

1. 일시 : 2003년 6월 27일 금요일 오후 4:00-6:00
2. 장소 : 한국방송통신대학교 세미나실
3. 참석자 : 이병직 서기관, 이화영 사무관, 천경기(노동부),
박천수(한국산업인력공단 중앙고용정보원 동향분석팀장)
이계오, 김영원, 이기재(연구진)

■ 주요 논의 사항

1. 사업체의 산업분류와 근로자 직종 분류에 대한 문제

임금구조통계조사에서 사업체에 대한 산업분류는 비교적 수월하고 정확하다. 세무서에 산업구분에 따라서 5 digits까지 신고하고 있다. 산업분류의 변화는 생산품의 변화에 따라서 발생할 수 있지만 직종분류에 비해서 상대적으로 수월하게 파악할 수 있다.

근로자를 대상으로 하는 조사에서 직종을 정확히 구분하는 것은 대단히 중요하다. 근로자의 직종 분류는 데이터의 수집도 어렵고, 비용도 많이 소요된다. 현재 노동부에서 사용하고 있는 한국표준직업분류는 직종 분류와 코딩 과정이 어렵다. 이에 비해서 중앙고용정보원에서 사용하고 있는 WIC-OES에 의한 직종분류 방법은 상대적으로 직종 분류와 코딩 과정이 편리하다는 장점이 있다. 또한 WIC-OES에 의해서 이루어진 분류는 한국표준직업분류에 의한 분류형태로 변환될 수 있다. 이런 점을 고려할 때 향후 근로자의 직종을 조사할 때 WIC-OES를 사용하는 방안을 검토할 수 있을 것이다.

2. 노동력수요동향조사에 대해서

노동력수요동향조사 통계는 사업체를 대상으로 산업소분류와 직종세분류(4 digits)를 기준으로 작성하고 있다. 따라서 이 조사의 결과는 근로자 수를 파악하는데 중요한 보조정보로 활용될 수 있을 것이다. 새로운 표본설계는 소영역통계 작성을 염두에 두고 이루어져야 할 것이다.

3. 임금구조통계조사의 목적에 대해서

임금구조통계조사의 가장 중요한 조사 목적은 임금구조의 파악이다. 따라서 소지역통계 작성기법을 적용하는 것도 직종세분류별 임금 구조를 파악하는 데 주안점을 두어야 할 것이다.

4. 직업분류 세분화에 따른 응답부담 가중

노동부의 입장은 근로자 개개인의 직종을 4자리까지 구분해야 하기 때문에 조사부담이 가중될 것이다. 2002년도에는 대규모 사업체인 경우에 해당 사업체에서 직접 근로자에 대한 직종을 자계식으로 기입하도록 하였다. 올해부터 직종을 4자리까지 구분해야 하기 때문에 자계식 조사를 실시하는 경우에 분류오류가 상당할 것으로 생각된다. 따라서 직종분류를 정확하고, 실무에서 쉽게 적용할 수 있는 방안으로 하는 것이 중요하다. 또한 표본 사업체의 응답부담을 줄여주기 위하여 사업체 내의 근로자 추출률을 다소 낮추는 방안을 함께 검토해야 한다.

5. 조사데이터의 질 관리에 대해서

근로자를 대상으로 하는 조사에서 근로자의 직종을 분류하는 문제는 대단히 어렵고 비용이 많이 소요된다. 조사 데이터의 질을 관리하기 위해서는 전체적으로 조사 근로자 수를 줄이고 좀더 정확한 조사가 될 수 있도록 노력하는 것이 바람직할 것이다. 노동부에서도 전체적인 노동통계 작성을 개편하여 사업체의 조사응답 부담을 덜어주기 위하여 다각적인 연구를 진행할 예정이다.

6. 중앙고용정보원의 고용구조조사와 임금구조통계조사의 관련성

중앙고용정보원의 고용구조조사는 가구를 대상으로 하는 취업자에 대한 조사이다. 따라서 사업체를 대상으로 하는 임금구조통계조사와 고용구조조사는 상호 보완적인 관계에 있다. 두 조사의 결과를 종합하여 우리 나라 전체의 고용구조를 파악하는 것이 가능할 것이다.

7. 우리 나라 노동통계 생산 전반에 걸친 내용

노동통계에 대한 패러다임의 전환이 필요한 시점이다. 과거에 상용근로자를 중심으로 작성되어 왔던 노동통계는 이제 정규근로자와 비정규근로자 구분으로 전환되었다. 또한 근로자의 직업분류도 우리의 실정에 맞는 새로운 방법으로 이루어져야 한다는 논의도 활발하다. 이러한 관점에서 보면 과거 노동통계 시계열과의 단절이 필요할 수도 있다. 아울러 표본 사업체들의 응답부담을 줄여주기 위한 다각적이 노력이 필요하다. 표본사업체들이 서로 중복을 피할 수 있도록 해야 하고, 사업체 내의 조사 근로자 추출률을 줄이는 방안도 고려해야 한다. 비슷한 유형의 노동통계조사를 통합해서 실시하는 방안도 고려해야 한다.

8. 기타

- 각종 임금관련 통계작성에 중앙값을 이용하는 것이 바람직하다. 다만, 당분간은 평균과 중앙값을 병기하는 방안을 모색해야 할 것이다.

- 연구제목에서 소지역이라는 용어는 연구 내용을 볼 때 부적절한 표현으로 사용자의 혼란을 초래할 수 있을 것으로 생각된다. 소지역이라는 용어는 다소는 지역적인 개념이 우선하는 것으로 보인다. 소영역 또는 소단위 통계 작성기법이라는 표현이 좀더 적절할 것이다.

(2) 2차 중간연구협회의 결과 정리

1. 일시 : 2003년 10월 21일 오후 4:30-6:30
2. 장소 : 한국방송통신대학교 전산센터 회의실
3. 참석자 : 이병직 서기관, 이화영 사무관, 천경기(노동부 노동경제담당관실)
박천수 박사(중앙고용정보원 동향분석팀장),
이계오, 김영원, 이기재(연구진)

1. 소지역 통계기법을 적용할 범위에 대한 논의

임금에 대한 소지역 통계는 정규근로자를 대상으로 한다. 비정규근로자를 포함해서 전체 근로자를 대상으로 접근하면 정규근로자와 비정규근로자의 여러 특성 차이로 인하여 적용상에 어려움이 많을 것이다.

2. 통계작성 범위와 방법에 대한 논의

모집단 분석을 통해서 또는 기존 자료를 분석해서 어느 범위까지 소지역 통계를 작성할 것인지 결정해야 한다. 산업소분류별 통계 작성을 위해서는 표본 사업체 수가 작은 경우에 구체적인 처리 원칙을 연구해야 한다. 이 경우에 고려할 수 있는 것은 산업소분류별 표본 사업체 수와 모집단 사업체 수를 분석하여 추출률을 계산하여 참고할 수 있다. 직종별 세분류별 통계를 위해서는 직종세분류별 조사 근로자 수가 작은 직종은 분석에서 제외하는 것을 원칙으로 한다. 중앙고용정보원의 직업지도(job map)에서는 조사 근로자 수가 30명 미만인 직업에 대해서는 통계를 작성하고 있지 않다. 다른 방안으로 우선 조사 데이터에 나타나는 모든 산업소분류와 직종세분류에 대해서 통계를 작성하고, 표본이 적거나 추정의 오차가 큰 경우에 대해서 이를 결과표에 표시하도록 하여 사용자가 사용하는데 도움을 줄 수 있을 것이다.

3. 산업소분류의 군집화 방안에 대해서

소지역 통계작성은 결국 비슷한 유형의 산업소분류들을 묶어서 합성추정량을 만들어서 직접추정량과 결합하여 사용하는 과정으로 볼 수 있다. 이를 위해서는 우선적으로 유사한 산업소분류들을 정의해서 묶는 작업이 필요하다. 서로 유사한 산업소분류들을 군집화는 우선적으로 산업별 특성을 고려해야 한다. 참고로 중앙고용정보원의 job map 정보를 이용할 수 있을 것이다.

4. 직종세분류별 소지역 통계 작성에 대한 논의

본 연구에서 가장 중요한 부분은 직종세분류별 임금통계 작성이다. 현재 분석 중인 2002년임금구조기본통계조사는 직종소분류로 조사되었기 때문에 자료 특성상 직종세분류별 임금통계 작성에

대한 연구가 어려운 형편이다. 그러나 직종세분류별 임금통계 작성이 연구의 핵심이라는 점을 고려하면 직종세분류별 임금통계 작성에 대한 구체적인 연구가 필요하다. 직종소분류별 임금통계 작성에 대한 구체적인 연구를 위해서는 중앙고용정보원의 고용구조조사를 이용할 수 있을 것이다.

5. 임금통계에 중앙값 사용에 대해서

임금통계에 대한 연구를 진행하면서 임금의 중앙값 작성에 대해서도 검토하는 것이 필요할 것이다. 소지역통계 분야에서 중앙값에 대한 추정 은 새로운 연구로 볼 수 있다. 이 부분에 대해서 추가적인 연구가 필요할 것이다.

6. 외부 보조정보 이용방안에 대한 논의

산업별 종사자 수, 직종별 종사자 수, 성별 구성 비율 등에 대한 정보는 노동력수요동향조사 결과를 이용할 수 있을 것이다. 노동력수요동향조사 결과를 이용하기 위해서는 산업별 종사자 수, 직종별 종사자 수 등에 대한 통계적 신뢰성을 확인하는 작업이 선행되어야 할 것이다. 다른 외부 정보로는 고용보험 DB 자료와 건강보험 자료 등을 이용할 수 있을 것이다. 그런데 고용보험 DB는 임금이 대한 정보가 갱신되지 않는다는 단점이 있고, 건강보험 자료는 최근의 임금을 파악할 수 있지만 전체 사업체를 포괄하고 있지 못하다. 다양한 외부 보조정보 자료에 대한 추가적인 검토작업이 필요하다.

7. 고용구조조사 이용방안에 대해서

고용구조조사는 기본적으로 가구를 대상으로 하는 통계조사로 사업체를 대상으로 하는 노동부의 통계조사와는 차이가 있다. 그러나 가구조사와 사업체 조사는 서로 상호보완적인 관계라는 점을 고려하면 고용구조조사 결과가 소지역통계 작성 결과의 타당성을 확인하는 근거자료로 이용될 수 있을 것이다.

8. 기타 논의사항

- 외국의 노동통계와 관련된 소지역통계 적용 사례를 정리하는 것이 좋겠다.
- 통계분석에서 사용되는 임금은 임금총액을 기준으로 한다.
- 복합추정량의 가중값을 결정할 때는 이론적 최적 가중값과 표본크기 의존 가중값을

모두 적용해서 현실적으로 적용 가능한 방법을 선택해야 할 것이다.

제 2 장 소지역 추정법(Small Area Estimation)

세부 단위(영역)에 배정된 표본수가 적을 경우에는 특성이 유사한 인근의 세부단위들을 결합하여 세부단위들의 그룹을 만들고 그 그룹 내에서 세부단위들은 동일한 특성을 갖는다고 가정할 수 있으면 유사한 인근의 조사결과를 활용해서 좀더 精度 높은 추정값을 작성할 수 있다. 이와 같이 비슷한 특성을 갖는 인근의 조사결과나 행정업무자료 또는 센서스 등 다른 통계조사 정보를 조사된 자료에서 얻은 직접 추정값과 결합하여 세부단위에 대한 통계를 작성하는 기법을 소지역 추정법(Small Area Estimation)이라고 한다. 일반적으로 소지역 추정법은 크게 직접추정법, 간접추정법과 모형기반 추정법으로 나눌 수 있다. 이 장에서는 소지역 추정법의 개념을 정리하고, 미국이나 캐나다와 같은 통계 선진국에서 적용하고 있는 다양한 소지역 추정법들과 노동통계에서 적용되는 소지역 추정법을 사례중심으로 살펴본다.

2.1 직접 추정법(Direct Estimation)

직접 추정법은 표본조사에서 해당 통계생산단위에 배정되어 조사된 자료들만을 이용하여 해당 통계생산단위의 추정치를 계산하는 방식을 말한다. i 통계생산 단위의 모수에 대한 추정량은 조사 모집단과 표본조사간의 구조관계에서 산출한 승수 w_j 와 관찰된 자료 y 들의 일차결합으로 다음과 같이 표현할 수 있다.

$$\hat{Y}_i = \sum_{j \in s_i} w_j y_j \quad (2.1)$$

여기에서 s_i 는 i 통계생산 단위에서 조사한 조사단위들의 집합이다.

식 (2.1)에서 주어진 추정량은 불편추정량이 되도록 w_j 를 산정하나 해당 통계생산 영역내의 조사단위가 적은 경우에는 추정의 精度가 떨어지고, 통계생산 영역에 따라서 표본 크기의 불균형적 분포로 추정량의 분산에 차이가 크게 발생하여 추정량의 신뢰성 창로 인해서 통계이용에 문제가 있을 수 있다.

2.2 간접 추정법(Indirect Estimation)

2.2.1 인구통계적 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 관련

된 징후 변수(출생자수, 사망자수, 주택 수, 등록된 학생 수 등)의 변동을 분석하여 얻은 예측값을 결합하는 추정법을 인구 통계학적 방법이라 말한다.

(1) 생존률법(Vital Rates Method: VR Method)

VR법은 출생과 사망에 관련된 자료를 이용하여 인구의 변동률보다는 징후 변수의 영향만을 분석하여 활용한다. 가장 최근에 센서스를 실시한 해를 기준 년도로 하고, 기준해로부터 t 년 후에 소지역의 인구수를 추정하고자 한다. 여기에서 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 y_{bt} 와 y_{dt} 로 표현하고 대지역의 출생률과 사망률을 R_{bt} 와 R_{dt} 라 나타내면 다음과 같은 관계가 주어진다.

$$y_{bt} = y_{bo} \left(\frac{R_{bt}}{R_{bo}} \right), \quad y_{dt} = y_{do} \left(\frac{R_{dt}}{R_{do}} \right) \quad (2.2)$$

여기에서 y_{bo} 와 y_{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 인구수는 다음 식에 의해서 추정할 수 있다.

$$p_t = \frac{1}{2} \left(\frac{b_t}{y_{bt}} + \frac{d_t}{y_{dt}} \right) \quad (2.3)$$

단, b_t 와 d_t 는 소지역의 t 년 후의 출생자수와 사망자수를 뜻한다.

(2) 성분법(Components 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 각

각 $b_{0,t}$, $d_{0,t}$, $m_{0,t}$ 로 나타냈을 때 t 년 후의 인구수는 다음 식에 의해 추정한다.

$$p_t = p_0 + b_{0,t} - d_{0,t} + m_{0,t} \quad (2.4)$$

여기에서 $m_{0,t} = i_{0,t} - e_{0,t} + n_{0,t}$ 로 계산하며, $i_{0,t}$ 는 유입인구, $e_{0,t}$ 는 유출인구, $n_{0,t}$ 는 주 간의 총 이주인구를 나타내며 행정보고자료에 의해 주어진다.

(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.

$$Y_i = (p_{it}/P_t) / (p_{i0}/P_0) = i \text{ 소지역의 인구비 변화량,}$$

$$x_{ij} = (s_{ijt}/S_{jt}) / (s_{ij0}/S_{j0}) = i \text{ 소지역에 대한 } j \text{ 번째 징후변수 } s_j \text{의 변화량,}$$

여기에서 P_t , P_0 , S_{jt} , S_{j0} 는 소지역 i 를 포함하는 넓은 지역에서의 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 일차결함으로 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 표본 회귀추정량을 이용한다.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}, \quad i=1, 2, \dots, m \quad (2.5)$$

i 소지역에 대한 인구수는 (2.5)식의 표본 회귀추정량을 이용하여 다음 식으로 추정한다.

$$\tilde{p}_{it} = \hat{Y}_i (p_{i0}/P_0) \hat{P}_t, \quad i=1, 2, \dots, m \quad (2.6)$$

표본 회귀추정량은 표본으로부터 직접 추정된 값이 아니라 다중선형회귀를 거쳐 얻어진 보정된 추정량이며, 표본 회귀법은 이를 이용하여 소지역의 인구를 추정하는 방법이다. 그러나 이러한 방법은 추후 논의될 모형에 근거한 소지역 추정정보는 효율성이 상당히 떨어지는 것으로 밝혀지고 있다.

2.2.2 합성 추정법(Synthetic Estimation)

추정하고자 하는 소지역과 특성이 유사한 소지역들의 정보를 이용하여 추정값의 정도를 높이고자하는 추정방식을 합성 추정법이라하며, 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. 우리나라의 경찰조사의 예로 살펴보면 조사단위는 동부와 읍면부로 나누어져 광역단체별로 층화추출법으로 선정되므로 시 지역과 읍면 지역으로 그룹을 나눈다면 각 그룹내에서는 연령대별 구조나 교육정도별 구조의 특성이 거의 유사할 것으로 판단할 수 있고, 이러한 경우 시군구 실업자 추정에 합성추정법의 이용도 가능할 것으로 생각된다.

대영역을 I 개 소지역으로 분할하며 또한 대영역을 특성 기준에 따라 J 개의 범주로 분류한다면 i 소지역의 추정값은 다음 식으로 구할 수 있다.

$$\hat{Y}_i = \sum_j^J p_{ij} \hat{Y}_{.j} \quad (2.7)$$

단, p_{ij} 는 i 번째 소지역의 j 범주에 대한 가중값이며 센서스나 행정자료에서 구해진다. $Y_{.j}$ 는 대영역에서 j 범주에 대한 표본에서 구한 추정값이다. 단 대영역의 표본의 수는 충분히 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,

$\hat{Y}_{.j} = \sum_i Y_{ij}$: j 범주의 대영역에 대한 합계,

$Y_i = \sum_j Y_{ij}$: i 소지역의 실업자 수.

$\hat{Y}_{.j}$ 의 직접 추정값 $\hat{Y}_{d,j}$ 는 표본조사 자료만으로 추정가능하고, X_{ij} 는 센서스 또는 행정자료 등 보조변수의 정보에서 계산 가능한 것으로 가정한다면 합성추정량은 다음과 같이 나타낼 수 있다.

$$\hat{Y}_i^s = \sum_j \left(\frac{X_{ij}}{X_{.j}} \right) \hat{Y}_{d,j} \quad (2.8)$$

만일 $\hat{Y}_{d,j}$ 가 비 추정량의 형식을 갖는다면, $\hat{Y}_{d,j} = (\hat{Y}_{.j} / \hat{X}_{.j}) X_{.j}$ 로 나타낼 수 있으므로 (2.8)식은 다음과 같이 표현될 수 있다.

$$\hat{Y}_i^s = \sum_j X_{ij} \left(\frac{\hat{Y}_{.j}}{\hat{X}_{.j}} \right) = \sum_j \left(\frac{X_{ij}}{\hat{X}_{.j}} \right) \hat{Y}_{.j} \quad (2.9)$$

여기에서 \hat{Y}_i^s 가 불편추정량이 되기 위해서는 $\frac{Y_{.j}}{X_{.j}} = \frac{Y_{ij}}{X_{ij}}$ 를 만족해야 하고, 이를 만족하

지 못할 경우에는 편향추정량이 되고, 이때 \hat{Y}_i^s 의 편향(Bias)의 크기는

$B(\hat{Y}_i^s) = E(\hat{Y}_i^s - Y_i)$ 이다. 즉, $B(\hat{Y}_i^s) = \sum_j X_{ij} \left(\frac{Y_{.j}}{\hat{X}_{.j}} - \frac{Y_{ij}}{X_{ij}} \right)$.

\hat{Y}_i^s 의 평균제곱오차($MSE(\hat{Y}_i^s)$)의 근사적 불편추정량이 아래와 같이 주어질 수 있다.

$$\widehat{MSE}(\hat{Y}_i^s) = (\hat{Y}_i^s - \hat{Y}_i)^2 - \widehat{Var}(\hat{Y}_i) \quad (2.10)$$

2.2.3 복합 추정법(Composite Estimation)

소지역에 배정된 표본수가 적기 때문에 표본 조사만을 이용한 직접 추정량의 불안정

에서 오는 낮은 신뢰성과 합성추정량의 편향을 보완하기 위해서 직접 추정값과 합성 추정값의 가중평균은 사용하는데 이를 복합 추정량(Composite Estimator)이라 한다.

$$\hat{Y}_i^c = w_i \hat{Y}_i + (1 - w_i) \hat{Y}_i^s \quad (2.11)$$

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.

$$w_{i(opt)} = \frac{MSE(\hat{Y}_i^s)}{MSE(\hat{Y}_i^s) + V(\hat{Y}_i)} \quad (2.12)$$

최적 가중값 $w_{i(opt)}$ 의 추정값은 다음 식으로 계산된다.

$$\hat{w}_{(opt)} = \frac{mse(\hat{Y}_i^s)}{(\hat{Y}_i^s - \hat{Y}_i)^2} \quad (2.13)$$

모든 소지역에 공통 가중값을 부여하는 방법으로써 초기 공통 가중값 w 를 이용하여 $MSE(\hat{Y}_i^s)$ 들의 평균을 최소화하는 가중값은 아래와 같다.

$$\hat{w}_{(opt)} = 1 - \frac{\sum_i \hat{V}(\hat{Y}_i)}{\sum_i (\hat{Y}_i^s - \hat{Y}_i)^2} \quad (2.14)$$

각 소지역에 배정된 표본 크기에 의존하는 가중값은 다음과 같이 계산된다.

$$w_i(\delta) = \begin{cases} 1, & \hat{N}_i \geq \delta N_i \\ \frac{\hat{N}_i}{\delta N_i}, & (\text{그외}) \end{cases} \quad (2.15)$$

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 실업자수와 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 생존률법(VR Method), 합성추정법 또는 복합 추정법 중 어느 한 방법으로 계산한 것으로 간주할 때 조정된 소지역 추정량은 다음과 같다.

$$\hat{Y}_i^A = \left(\frac{\hat{Y}_i^*}{\sum_I \hat{Y}_i^*} \right) \hat{Y} \quad (2.16)$$

단, \hat{Y} 는 광역시·도의 직접 추정값이며, \hat{Y}_i^* 는 i 소지역의 *추정법으로 추정한 것이다.

2.3 모형 기반 추정법(Model-Based Estimation)

2.3.1 기본적인 지역 수준 모형(Basic Area-level Model)

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 있으며, 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 최량선형불편 예측(EBLUP) 추정량, 경험적 베이스(EB) 추정량, 계층적 베이스(HB) 추정량에 대해 설명하기로 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.

$$\widehat{\Theta}_i = \Theta_i + e_i, \quad i=1,2,\dots,m \quad (2.17)$$

여기에서 표본오차 e_i 는 서로 독립이며, 평균이 0, 분산이 ψ_i 임이 가정되며, 보통 ψ_i 는 알려진 것으로 가정한다.

Θ_i 는 소지역의 정보를 나타내는 보조변수 $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{ip})^T$ 를 이용하여 선형 회귀모형을 통해 표현한다.

$$\begin{aligned} \Theta_i &= z_{i1}\beta_1 + z_{i2}\beta_2 + \dots + z_{ip}\beta_p + v_i \\ &= \mathbf{z}_i^T \boldsymbol{\beta} + v_i \end{aligned} \quad (2.18)$$

여기에서 모형오차 v_i 는 서로 독립이며, 평균이 0, 분산 σ_v^2 을 갖고, 표본오차 e_i 와는 서로 독립임을 가정한다.

마지막으로 (2.17)과 (2.18)의 두 성분들을 결합하면 다음과 같은 결합모형을 얻을 수 있다.

$$\widehat{\Theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i + e_i \quad (2.19)$$

위의 결합모형은 고정효과 $\boldsymbol{\beta}$ 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 척도이다.

2.3.2 경험적 최량선형불편예측(EBLUP) 방법

경험적 최량선형불편예측(EBLUP) 방법, 경험적 베이지(EB) 방법 및 계층적 베이지(HB) 방법은 모형에 근거한 소지역 추정문제에 많이 활용되고 있는 방법이다. 특히 경험적 최량선형불편예측 방법은 선형혼합모형을 이용한 추론에 응용되어 왔고, 경험적 베이지 방법 및 계층적 베이지 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, MSE추정을 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 변동성을 나타내는 측도들은 동일하지는 않다.

고정계수 I_t 를 갖는 θ_i 의 선형추정량 $\sum I_t \hat{\theta}_t$ 가 모형 (4.3)에 대해서 $\sum I_t \hat{\theta}_t - \theta_i$ 의 기대값이 0을 만족할 때, $\sum I_t \hat{\theta}_t$ 를 θ_i 의 선형불편예측(LUP) 추정량이라 한다. θ_i 의 최량선형불편예측(BLUP) 추정량은 선형불편예측(LUP) 추정량들 중 최소평균제곱오차를 갖는 추정량을 말한다.

모형 (2.19)하에서 θ_i 의 BLUP 추정량은 다음과 같이 주어진다(Prasad and Rao,1990).

$$\bar{\theta}_i(\sigma_v^2) = y_i \hat{\theta}_i + (1 - y_i) \mathbf{z}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2) \quad (2.20)$$

여기에서 $y_i = \sigma_v^2 / (\sigma_v^2 + \psi_i)$ 이고, $\tilde{\boldsymbol{\beta}}(\sigma_v^2)$ 은 가중치 $(\sigma_v^2 + \psi_i)^{-1}$ 을 갖는 가중최소제곱 추정량으로 아래와 같이 주어진다.

$$\tilde{\boldsymbol{\beta}}(\sigma_v^2) = (\sum_I y_i \mathbf{z}_i \mathbf{z}_i^T)^{-1} (\sum_I y_i \mathbf{z}_i y_i) \quad (2.21)$$

(2.20)식의 BLUP 추정량은 가중치 y_i 를 갖는 직접추정량 $\hat{\theta}_i$ 과 가중치 $1 - y_i$ 를 갖는 회귀합성추정량 $\mathbf{z}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2)$ 의 가중결합으로 볼 수 있다. 또한, 표본분산 ψ_i 가 작을 때 (σ_v^2 이 클 경우) BLUP 추정량은 직접추정량 $\hat{\theta}_i$ 에 큰 가중치가 부여되고, 반대의 경우에는 회귀합성추정량 $\mathbf{z}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2)$ 에 큰 가중치가 부여된다. 표본이 추출되지 않은 지역들에 대해서는 BLUP 추정량은 회귀합성추정량으로 주어질 수 있다.

BLUP 추정량의 변동성의 측도는 추정량의 $MSE = E(est. - \theta_i)^2$ 에 의해 주어지며 다음과 같다.

$$MSE\{\bar{\theta}_i(\sigma_v^2)\} = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) \quad (2.22)$$

단, $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$ 이고, $g_{2i}(\sigma_v^2) = \sigma_v^2 (1 - \gamma_i)^2 \mathbf{z}_i^T (\sum_i \gamma_i \mathbf{z}_i \mathbf{z}_i^T)^{-1} \mathbf{z}_i$ 로 주어진다. 식 (2.20)와 (2.22)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 $O(m^{-1})$ 의 형식 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 표본분산에 비해 모형분산을 작게할 수 있는 보조변수에 크게 의존한다고 볼 수 있다.

대부분의 문제에서는 모형분산 σ_v^2 은 미지인 값이므로 적절한 $\widehat{\sigma}_v^2$ 을 추정하여 EBLUP 추정량 $\bar{\Theta}_i = \bar{\Theta}_i(\widehat{\sigma}_v^2)$ 을 산출한다. 이때 소지역의 평균 \bar{Y}_i 의 추정량은 $g^{-1}(\bar{\Theta}_i)$ 로, σ_v^2 의 추정량은 $\widehat{\sigma}_v^2 = \max(\widetilde{\sigma}_v^2, 0)$ 로 주어진다. 여기에서 $\widetilde{\sigma}_v^2$ 은 다음 식을 만족한다.

$$(m-p) \widetilde{\sigma}_v^2 = \sum_i (\widehat{\Theta}_i - \mathbf{z}_i^T \boldsymbol{\beta}^*)^2 - \sum_i \psi_i h_{ii} \quad (2.23)$$

(2.23)식에서 $h_{ii} = \mathbf{z}_i^T (\sum_i \mathbf{z}_i \mathbf{z}_i^T)^{-1} \mathbf{z}_i$ 이고, $\boldsymbol{\beta}^*$ 는 $\boldsymbol{\beta}$ 의 OLS(ordinary least squares) 추정량이다. 한편, $\widetilde{\sigma}_v^2$ 은 다음과 같은 비선형 방정식의 반복적인 해로써 구할 수도 있다.

$$a(\sigma_v^2) = \sum_i \{ \widehat{\Theta}_i - \mathbf{z}_i^T \widetilde{\boldsymbol{\beta}}(\sigma_v^2) \}^2 / (\sigma_v^2 + \psi_i) = m-p \quad (2.24)$$

여기에서 $\widetilde{\boldsymbol{\beta}}(\sigma_v^2)$ 은 (2.21)식에 주어졌고, (2.24)식의 가운데 항은 가중잔차제곱합, $m-p$ 는 가중잔차제곱합과 관계가 되는 자유도이다. 만약 $\widehat{\sigma}_v^2 = 0$ 이면 EBLUP 추정량 $\bar{\Theta}_i$ 는 회귀합성추정량 $\mathbf{z}_i^T \widehat{\boldsymbol{\beta}}$ 로 축약된다. 단, $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\sigma}_v^2)$ 이며, 식(2.21)에서 σ_v^2 대신에 $\widehat{\sigma}_v^2$ 을 대체하여 산출한다. 물론 위의 (2.23), (2.24)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, $\widehat{\theta}_i$ 는 평균이 $\mathbf{z}_i^T \boldsymbol{\beta}$ 이고 분산이 $\sigma_v^2 + \psi_i$ 인 서로 독립인 정규분포를 따르게 된다. 이러한 분포에 대한 가정 하에서 계산된 $\boldsymbol{\beta}$ 와 σ_v^2 의 최대우도추정량을 제한최대우도추정량(REML)이라 하며, 선형혼합모형하에서도 근사적으로 유효하다. 따라서 $\bar{\theta}_i$ 의 BLUP 추정량을 산출 시 σ_v^2 의 REML 추정량을 이용하여도 근사적으로 타당하다.

2.3.3 경험적 베이즈(EB) 방법

경험적 베이즈(EB) 추정법은 랜덤오차 v_i 와 e_i 가 정규분포를 따른다는 가정하에서 출발한다. $(\widehat{\theta}_i, \theta_i)$ 의 결합분포가 평균이 $(\mathbf{z}_i^T \boldsymbol{\beta}, \mathbf{z}_i^T \boldsymbol{\beta})$, 분산이 $(\sigma_v^2 + \psi_i, \sigma_v^2)$, 상관 계수가 γ_i 인 이변량 정규분포를 따른다고 가정하자. 이때, θ_i 의 평균제곱오차를 최소화하는 베이즈 추정량은 다음과 같다.

$$\bar{\theta}_i^B(\boldsymbol{\beta}, \sigma_v^2) = E(\theta_i | \widehat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2) = \gamma_i \widehat{\theta}_i + (1 - \gamma_i) \mathbf{z}_i^T \boldsymbol{\beta} \quad (2.25)$$

(2.25)식의 베이즈 추정량은 선형성 또는 불편성을 만족하지는 않는다. 여기에서 모수 $\boldsymbol{\beta}$ 와 σ_v^2 을 제한최대우도(REML) 추정량으로 대체하여 다음과 같은 θ_i 의 경험적 베이즈(EB) 추정량을 얻는다.

$$\bar{\theta}_i^{EB} = \bar{\theta}_i^B(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_v^2) \quad (2.26)$$

경험적 베이즈(EB) 추정량 $\bar{\theta}_i^{EB}$ 는 정규분포의 가정하에서는 EBLUP 추정량 $\bar{\theta}_i$ 와 같다. 그러나 경험적 베이즈방법은 $\widehat{\theta}_i$ 과 θ_i 의 임의의 결합분포에 대해서도 일반적으로 응용할 수 있다는 점을 장점으로 들 수 있다.

EBLUP 추정량 $\bar{\theta}_i = \bar{\theta}_i(\widehat{\sigma}_v^2)$ 의 MSE 추정량은 (4.6)식에서 σ_v^2 대신 $\widehat{\sigma}_v^2$ 을 대체하여 얻어질 수 있으나, 이 경우에는 σ_v^2 에 대한 추정효과가 무시되기 때문에 MSE의 추정값은 과소추정되는 경향을 보인다. 이러한 문제 때문에 Prasad and Rao(1990)는

$\{v_i\}$ 와 $\{e_i\}$ 에 대해 정규성을 가정하여 근사적으로 불편인 EBLUP 추정량 $\bar{\theta}_i$ 의 MSE 추정량을 제안하였다. Prasad and Rao(1990)가 제안한 MSE 추정량은 (2.23)식의 σ_v^2 의 적률추정량을 사용하였을 경우 다음과 같이 주어진다.

$$mse(\bar{\theta}_i) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) \quad (2.27)$$

단, $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$, $g_{2i}(\sigma_v^2) = \sigma_v^2 (1 - \gamma_i)^2 \mathbf{z}_i^T (\sum_I \gamma_i \mathbf{z}_i \mathbf{z}_i^T)^{-1} \mathbf{z}_i$,

$g_{3i}(\sigma_v^2) = \{ \psi_i^2 / (\sigma_v^2 + \psi_i)^3 \} h(\sigma_v^2)$, $h(\sigma_v^2) = 2m^{-2} \sum_I (\sigma_v^2 + \psi_i)^2$ 로 주어진다.

최근들어 Jing, Lahiri and Wan(1999)는 근사적으로 불편인 켄나이프 MSE 추정량을 제안하였다. 켄나이프 방법은 랜덤인 지역효과들을 갖는 로지스틱 회귀와 같은 좀 더 복잡한 모형들에 대해서도 쉽게 적용할 수 있다는 장점을 갖고 있다.

θ_i 의 EB 추정량 (4.10)을 $\bar{\theta}_i^{EB} = k(\hat{\theta}_i, \square)$ 로 표현할 때, 켄나이프 절차는 다음과 같다. 여기에서 $\square = (\boldsymbol{\beta}, \sigma_v^2)$ 은 모형에서의 모수 $\boldsymbol{\beta}$ 와 σ_v^2 을 나타낸다.

(i) l 번째 지역의 자료 $(\hat{\theta}_l, \mathbf{z}_l)$ 을 제외한 \square 의 추정량 $\square(l)$ 을 계산한다.

이때의 EB 추정량을 $\bar{\theta}_i^{EB}(l) = k(\hat{\theta}_i, \square(l))$ 로 나타내자.

(ii) $\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^m (\bar{\theta}_i^{EB}(l) - \bar{\theta}_i^{EB})$ 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) 끝으로 MSE의 켄나이프 추정량 $mse_j(\bar{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$ 를 계산한다.

\hat{M}_{1i} 은 \square 가 기지일 때 MSE에 대한 추정량이며, \hat{M}_{2i} 는 모형 모수 \square 를 추정할 때 추가적으로 발생하는 MSE에 대한 변화량을 추정한다.

2.3.4 계층적 베이즈(HB) 방법

계층적 베이즈(HB) 방법을 이용한 추론은 비교적 추론의 정확도가 높고, 복잡한 유형의 문제들에서도 최근에 개발된 MCMC(Markov Chain Monte Carlo)방법을 이용하여 해결할 수 있다. 깃스 샘플러도 이러한 방법의 일종이다. HB 방법에서는 모형 모수 $\square = (\boldsymbol{\beta}, \sigma_v^2)$ 뿐만 아니라 모집단의 값 θ_i 가 랜덤으로 간주되며, 모형 모수들에 대한 사전분포가 명시된다. θ_i 들에 관한 추론은 주변 사후분포에 의해 결정된다. 즉, 주어진 자료 $\{(\widehat{\theta}_i, \mathbf{z}_i), i=1, 2, \dots, m\}$ 에 대한 조건부 분포 $f(\theta_i | \widehat{\boldsymbol{\theta}})$ 에 의해 추론이 행해진다. 여기에서 $\widehat{\boldsymbol{\theta}}$ 은 직접추정값 $\widehat{\theta}_i$ 의 벡터이다. 특히 θ_i 는 사후분포의 평균 $E(\theta_i | \widehat{\boldsymbol{\theta}})$ 에 의해 추정되며, 추정량의 변동은 사후분포의 분산 $V(\theta_i | \widehat{\boldsymbol{\theta}})$ 에 의해 추정된다.

먼저 σ_v^2 이 기지인 상태를 가정하고 $\boldsymbol{\beta}$ 에 관한 사전분포를 배정하기로 한다. $\boldsymbol{\beta}$ 의 사전분포가 상수에 비례하고(i.e improper prior), v_i 와 e_i 가 정규분포를 따른다고 가정한다면, 이때 사후평균 $E(\theta_i | \widehat{\boldsymbol{\theta}}, \sigma_v^2)$ 은 (2.20)식의 BLUP 추정량 $\bar{\theta}_i(\sigma_v^2)$ 과 동일하다. 더욱이 사후분산 $V(\theta_i | \widehat{\boldsymbol{\theta}}, \sigma_v^2)$ 은 (2.22)식의 BLUP 추정량의 MSE와 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 경우에는 $\boldsymbol{\beta}$ 뿐만 아니라 σ_v^2 에 관한 사전분포를 고려해야 하며, 또한 서로 독립임을 가정하여 주변사후분포 $f(\sigma_v^2 | \widehat{\boldsymbol{\theta}})$ 을 이끌어 낸다. 만약 σ_v^2 에 관한 사전분포를 불완전(improper) 사전분포를 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 배정한다(여기에서 $G(a, b) : f(\tau_v) \propto \exp(-a\tau_v) \tau_v^{b-1}$). 주변사후분포 $f(\sigma_v^2 | \widehat{\boldsymbol{\theta}})$ 를 이용한 HB 추정량 $E(\theta_i | \widehat{\boldsymbol{\theta}})$ 은 다음 식과 같이 주어진다.

$$\bar{\theta}_i^{HB} = E(\theta_i | \widehat{\boldsymbol{\theta}}) = \int \bar{\theta}_i(\sigma_v^2) f(\sigma_v^2 | \widehat{\boldsymbol{\theta}}) d\sigma_v^2 \quad (2.28)$$

위의 (2.28)식을 $E_{\sigma_v^2 | \widehat{\boldsymbol{\theta}}} \{ \bar{\theta}_i(\sigma_v^2) \}$ 으로 표현하면, 사후분산 $V(\theta_i | \widehat{\boldsymbol{\theta}})$ 은 다음과 같다.

$$V(\theta_i | \hat{\Theta}) = E_{\sigma_v^2 | \hat{\Theta}} \{g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2)\} + V_{\sigma_v^2 | \hat{\Theta}} \{\bar{\Theta}_i(\sigma_v^2)\} \quad (2.29)$$

여기에서 $V_{\sigma_v^2 | \hat{\Theta}}$ 은 $f(\sigma_v^2 | \hat{\Theta})$ 에 관한 분산을 의미한다.

위에서 소개한 (2.28)와 (2.29)은 일차원 수치적분에 의해 계산된다. 좀 더 복잡한 모형에 대한 고차원 수치적분은 MCMC 방법을 이용하여 계산할 수 있다. (2.28)식으로부터 $\bar{\Theta}_i^{HB}$ 는 EBLUP(EB) 추정량 $\bar{\Theta}_i(\hat{\sigma}_v^2)$ 와 근사적으로 같다는 것을 알 수 있다.

깁스 샘플링은 위의 (2.28)와 (2.29)를 결정하기 위해 사용될 수 있는 일종의 MCMC 방법이다. 깁스 샘플링을 수행하기 위해서는 다음과 같은 깁스 조건부 분포들이 필요하다.

$$(i) \beta | \theta, \sigma_v^2, \hat{\Theta} \sim N_p((\sum z_i z_i^T)^{-1}(\sum z_i \theta_i), \sigma_v^2(\sum z_i z_i^T)^{-1})$$

$$(ii) \theta_i | \beta, \sigma_v^2, \hat{\Theta} \sim N(\bar{\Theta}_i^B(\beta, \sigma_v^2), g_{1i}(\sigma_v^2) = \Psi_i \Psi_i)$$

$$(iii) \tau_v = \sigma_v^{-2} | \beta, \theta, \hat{\Theta} \sim G(\tilde{a}, \tilde{b}), \quad \text{단, } \tilde{a} = \frac{1}{2} \sum (\theta_i - z_i^T)^2 + a,$$

$$\tilde{b} = \frac{m}{2} + b.$$

깁스 알고리즘은 다음 절차에 의해 이루어진다.

$$(a) \theta_i = \theta_i^{(0)}, \sigma_v^2 = \sigma_v^{2(0)} \text{을 초기값으로 하여 위의 (i)로부터 } \beta^{(1)} \text{ 계산}$$

$$(b) \beta = \beta^{(1)}, \sigma_v^2 = \sigma_v^{2(0)} \text{를 이용하여 위의 (ii)로부터 } \theta_i^{(1)}$$

($i=1, 2, \dots, m$) 을 계산

$$(c) \theta_i = \theta_i^{(1)} \text{과 } \beta = \beta^{(1)} \text{을 이용하여 위의 (iii)으로부터 } \sigma_v^{2(1)} \text{을 계산}$$

(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

$\{ \boldsymbol{\beta}^{(t+j)}, \sigma_v^{2(t+j)}, \theta_1^{(t+j)}, \dots, \theta_m^{(t+j)}; j=1, 2, \dots, J \}$ 을 $\boldsymbol{\beta}, \sigma_v^2, \theta_1, \dots, \theta_m$ 의 결합 사후분포로 얻은 표본으로 간주한다. 초기값은 보통 $\theta_i^{(0)} = \bar{\theta}_i^{EB}, \sigma_v^{2(0)} = \sigma_v^2$ 의 REML 추정량을 사용한다.

위에서 계산된 J 개의 표본을 이용하여 θ_i 의 사후평균, 사후분산을 다음과 같이 추정한다.

$$\begin{aligned} \bar{\theta}_i^{HB} &\approx \frac{1}{J} \sum_j \bar{\theta}_i(\sigma_v^{2(t+j)}) \\ &= \frac{1}{J} \sum_j \bar{\theta}_i(j) = \bar{\theta}_i(\cdot), \end{aligned} \quad (2.30)$$

$$V(\theta_i | \hat{\boldsymbol{\theta}}) \approx \frac{1}{J} \sum_j \{g_{1i}(\sigma_v^{2(t+j)}) + g_{2i}(\sigma_v^{2(t+j)})\} + \frac{1}{J} \sum_j \{ \bar{\theta}_i(j) - \bar{\theta}_i(\cdot) \}^2 \quad (2.31)$$

2.4 소지역 평균의 추정량

(1) μ_i 의 EBLUP 추정량

소지역의 특성치를 추정하기 위하여 먼저 다음과 같은 모형을 가정하기로 한다.

$$Y_i = X_i \beta_i + \varepsilon_i,$$

$$\beta_i = Z_i \boldsymbol{\gamma} + v_i, \quad i=1, 2, \dots, m \text{ (소지역 수)}$$

여기에서 N_i 는 i 번째 소지역의 모집단 크기를 나타낸다.

위의 모형에서 i 번째 소지역에 대한 평균의 기댓값은 다음의 (3.5)식과 같이 표현할 수 있다.

$$\begin{aligned} \mu_i &= \bar{X}_i^t \beta_i \\ &= \bar{X}_i^t (Z_i \boldsymbol{\gamma} + v_i) \end{aligned}$$

$$= \bar{X}_i^t Z_i \gamma + \bar{X}_i^t v_i, \quad (3.5)$$

여기에서 $\bar{X}_i = (1, \bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{ip})^t$ 는 i 번째 소지역에 대한 모집단 평균벡터를 나타낸다.

μ_i 의 EBLUP 추정량 $\hat{\mu}_i$ 은 다음의 (3.6)식과 같이 표현된다.

$$\hat{\mu}_i = \bar{X}_i^t Z_i \hat{\gamma} + \bar{X}_i^t \hat{v}_i, \quad (3.6)$$

단, $\hat{v}_i = \hat{\Omega} X_i^t \hat{V}_i^{-1} (Y_i - X_i Z_i \hat{\gamma})$, $\hat{V}_i^{-1} = \hat{\sigma}^{-2} I - \hat{\sigma}^{-4} X_i \hat{\Omega} \hat{G}_i^{-1} X_i^t$,

$\hat{G}_i^{-1} = (I + \hat{\sigma}^{-2} X_i^t X_i \hat{\Omega})^{-1}$ 이며, 여기에서 γ 와 θ 의 추정량 $\hat{\gamma}$ 와 $\hat{\theta}$ 는 앞서 계산된 RIGLS 추정량이 사용된다.

절편항만 랜덤이고, β 의 나머지 항들은 고정계수들인 Battese et al.(1981, 1988)이 제안한 μ_i 의 EBLUP 추정량 $\hat{\mu}_{i(RD)}$ 은 다음의 (3.7)식과 같다.

$$\hat{\mu}_{i(RD)} = \bar{X}_i^t \hat{\beta} + \hat{v}_{i0} \quad (3.7)$$

(2) $\hat{\mu}_i$ 의 MSE 근사와 추정식

다음의 사실을 이용하여 $\hat{\mu}_i$ 의 MSE 근사식을 유도한다.

- i) μ_i 의 REMLE 는 translation invariant (Kackar & Harville(1984)).
- ii) μ_i 의 RIGLSE 는 정규성 가정 하에서 REMLE와 동치(Goldstein(1989))

위의 사실을 이용한다면, μ_i 의 RIGLSE 인 $\hat{\mu}_i$ 도 translation invariant 이므로 다음 식이 성립하게 된다.

$$MSE(\hat{\mu}_i) = E(\hat{\mu}_i - \mu_i)^2$$

$$= E(\tilde{\mu}_i - \mu_i)^2 + E(\hat{\mu}_i - \tilde{\mu}_i)^2, \quad i = 1, 2, \dots, m$$

단, $\tilde{\mu}_i$ 는 μ_i 의 BLUP 추정량이며, 여기에서 첫 번째 항인 $E(\tilde{\mu}_i - \mu_i)^2$ 은 다음 (3.8) 식의 결과에서 계산된다.

$$\begin{aligned} MSE(\tilde{\mu}_i) &= E(\tilde{\mu}_i - \mu_i)^2 \\ &= \bar{X}_i^t (G_i^{-1})^t \Omega \bar{X}_i \\ &\quad + \sigma^2 \bar{X}_i (G_i^{-1})^t Z_i \left(\sum_{j=1}^m Z_j^t G_j^{-1} X_j^t X_j Z_j \right)^{-1} Z_i^t G_i^{-1} \bar{X}_i \\ &= T_1 + T_2, \end{aligned} \quad (3.8)$$

$$\text{단, } G_i = I + \sigma^2 X_i^t X_i \Omega$$

두 번째 항 $E(\hat{\mu}_i - \tilde{\mu}_i)^2$ 은 (5.9) 식과 같이 근사적인 계산식으로 주어진다.

$$\begin{aligned} E(\hat{\mu}_i - \tilde{\mu}_i)^2 &\approx T_3 \\ &= \text{tr} \left[\left(\frac{\partial d_i}{\partial \theta} \right) V \left(\frac{\partial d_i}{\partial \theta} \right)^t E \{ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^t \} \right] \end{aligned} \quad (3.9)$$

여기에서 $d_i = \bar{X}_i^t K_i (I \otimes \Omega) X_i^t V^{-1}$, $K_i = [0, \dots, 0, I, 0, \dots, 0]$ 는 $(p+1) \times (p+1)m$ 행렬로써 i 번째 소지역에서 $(p+1) \times (p+1)$ 단위행렬을 가지며, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s)$ 은 $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ 의 translation invariant 추정량으로써, $\theta_s = \sigma^2$, $\theta_k (k=1, 2, \dots, s-1)$ 는 Ω 의 서로다른 원소를 나타낸다.

위의 (3.9) 식에서 $E \{ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^t \}$ 를 REML 추정량 B의 점근적 공분산 행렬로 근사시킨 계산 결과식은 다음 (3.10) 식과 같다.

$$T_3 = \bar{X}_i^t (G_i^{-1}) \left(\sum_{j=1}^{s-1} \sum_{k=1}^{s-1} b_{jk} \Delta_j C_i \Delta_k^t \right) G_i^{-1} \bar{X}_i$$

$$-2 \bar{X}_i^t (G_i^{-1})^t \left(\sum_{j=1}^{s-1} b_{j,s} \Delta_j \right) R_i \Omega \bar{X}_i + b_{ss} \bar{X}_i^t \Omega S_i \Omega \bar{X}_i, \quad (3.10)$$

여기에서 $B = \Theta$ 의 REML 추정량, $b_{j,k} = B$ 의 jk 번째 원소, $b_{jk}^* = B^{-1}$ 의 jk 번째 원

소 $= \text{tr} \left(\sum_{i=1}^m P_i \frac{\partial V}{\partial \theta_j} P_i \frac{\partial V}{\partial \theta_k} \right)$ ($k=1, 2, \dots, s$) 이며, P_i 는 다음과 같이 표현된다.

$$P_i = V_i^{-1} - V_i^{-1} X_i Z_i \left(\sum_{i=1}^m Z_i^t X_i^t V_i^{-1} X_i Z_i \right)^{-1} Z_i^t X_i^t V_i^{-1},$$

$$\text{단, } C_i = \sigma^{-2} G_i^{-1} X_i^t X_i,$$

$$R_i = \sigma^{-4} G_i^{-2} X_i^t X_i,$$

$$S_i = \sigma^{-6} G_i^{-3} X_i^t X_i,$$

$$\Delta_k = \frac{\partial \Omega}{\partial \Omega_k} : (s-1) \times (s-1) \quad (k=1, 2, \dots, s-1).$$

위의 (3.8)식과 (3.10)식을 정리하면, $\hat{\mu}_i$ 의 MSE 근사식은 다음 (3.11)식과 같이 표현된다.

$$\begin{aligned} \text{MSE}(\hat{\mu}_i) &= E(\hat{\mu}_i - \mu_i)^2 \\ &= E(\tilde{\mu}_i - \mu_i)^2 + E(\hat{\mu}_i - \tilde{\mu}_i)^2, \quad i=1, 2, \dots, m \\ &\approx \left\{ \bar{X}_i^t (G_i^{-1})^t \Omega \bar{X}_i \right\} \\ &\quad + \left\{ \sigma^2 \bar{X}_i^t (G_i^{-1})^t Z_i \left(\sum_{i=1}^m Z_i^t G_i^{-1} X_i^t X_i Z_i \right)^{-1} Z_i^t G_i^{-1} \bar{X}_i \right\} \\ &\quad + \left\{ \bar{X}_i^t (G_i^{-1}) \left(\sum_{j=1}^{s-1} \sum_{k=1}^{s-1} b_{jk} \Delta_j C_i \Delta_k^t \right) G_i^{-1} \bar{X}_i \right. \\ &\quad \left. - 2 \bar{X}_i^t (G_i^{-1})^t \left(\sum_{j=1}^{s-1} b_{j,s} \Delta_j \right) R_i \Omega \bar{X}_i + b_{ss} \bar{X}_i^t \Omega S_i \Omega \bar{X}_i \right\} \\ &= T_1 + T_2 + T_3 \end{aligned} \quad (3.11)$$

한편, 위의 결과들을 유한 모집단에 적용할 경우에는 다음과 같이 보정 절차를 거쳐 표본으로 추출되지 않은 단위들의 영향력을 고려해 주어야 한다.

$$\widehat{\mu}_i^F = f_i \bar{y}_i + (\bar{X}_i - f_i \bar{x}_i)^t (Z_i \widehat{\gamma} + \widehat{v}_i), \quad (3.12)$$

단, $f_i = n_i/N_i$, $\bar{x}_i = (1, \bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{ip})^t$ 는 표본평균을 나타낸다.

$$\widehat{\mu}_i^F - \bar{Y} = (1 - f_i) [(\bar{X}_i^C)^t \{Z_i(\widehat{\gamma} - \gamma) + \widehat{v}_i - v_i - \bar{\varepsilon}_i^C\}], \quad (3.13)$$

여기에서 $\bar{X}_i^C = (1 - f_i)^{-1}(\bar{X}_i - f_i \bar{x}_i)$, $\bar{\varepsilon}_i^C = 'i$ 번째 소지역에서 표본으로 추출되지 않은 단위들에 대한 ε_{ij} 의 평균'을 나타낸다.

위의 (3.12)식과 (3.13)식으로부터 추정량의 MSE 를 계산하면 다음 (3.14)식과 같이 주어진다.

$$MSE(\widehat{\mu}_i^F) = (1 - f_i)^2 \{MSE^*(\widehat{\mu}_i) + N_i^{-1}(1 - f_i)^{-1}\sigma^2\}, \quad (3.14)$$

여기에서 $MSE^*(\widehat{\mu}_i)$ 는 (3.11)식에서 \bar{X}_i 를 \bar{X}_i^C 로 대체하여 구한다.

$\widehat{\mu}_i$ 의 MSE 추정식은 $\widehat{\mu}_i$ 의 MSE 근사식 (3.11)과 (3.14)식에서 σ^2 과 Ω 의 RIGLS 추정량을 이용할 수 있으며, RIGLS 추정량을 이용한 $\widehat{\mu}_i$ 의 MSE 추정식은 (3.15)와 (3.16)과 같이 주어진다.

$$\widehat{MSE}(\widehat{\mu}_i) = \widehat{T}_1 + \widehat{T}_2 + \widehat{T}_3, \quad (3.15)$$

$$\widehat{MSE}(\widehat{\mu}_i^F) = (1 - f_i)^2 \{ \widehat{MSE}^*(\widehat{\mu}_i) + N_i^{-1}(1 - f_i)^{-1} \widehat{\sigma}^2 \}. \quad (3.16)$$

2.5 외국의 소지역 추정법 적용 사례

2.5.1 미국 주단위와 지역단위의 취업통계

미국경상취업통계조사(CES : current employment statistics)는 다른 표본조사와 같이

표본오차와 비표본오차를 갖는다. 표본오차의 크기 또는 분산은 표본이 대표할 수 있는 조사대상의 범위와 표본의 크기에 직접적인 관계가 있다.

주단위와 지역단위의 취업통계의 추정치는 두 가지 방법으로 산출되는데 대부분의 추정치들은 표본조사를 기반으로 한 직접추정법으로 계산된다. 응답한 표본만으로 취업통계를 생산하기에 표본크기가 충분하지 않을 경우에는 소영역 모형(small domain model)을 이용하여 주단위와 지역단위의 통계를 생산하기도 한다.

(1) 취업통계 추정치의 분산(Variance of the CES estimates)

기술보고서(technical note)에 표기된 표본오차들은 표본조사를 기반으로 한 직접 추정치와 실직보험(UI : unemployment insurance)에서 획득한 과거의 취업정보를 이용하여 추정한 값들이다.

취업과 임금의 총괄 프로그램(ES-202)은 실직보험텍스보고서로부터 추론하여 취업자의 전체에 대한 거의 완벽한 명부를 제공한다. 실직보험자료는 표본추출틀과 CES추정치들의 연단위의 벤치마킹정보를 제공한다. 표본오차는 평균응답률과 지난 실직보험보고서에서 보고된 취업패턴으로부터 계산된다. 무응답률이 평균수준이고 이상한 응답값이 없는 정상적인 상황에서 표본조사자료만으로 계산한 추정치들의 오차의 크기를 나타내는 좋은 지표이다. 표본추출 이외의 이유에서 발생하는 오차는 비표본오차로서 벤치마킹의 수정을 통해서 알아낼 수 있다.

(2) CES 벤치마킹 수정(The CES benchmark revision)

표본오차와 비표본오차의 조정을 위해서 매년 CES 추정치를 취업자 총계에 벤치마킹한다. 취업자의 총계는 대부분의 미국내의 취업자들은 미국취업안전원(State Employment Security Agencies)에 보고하도록 되어있는 실직보험텍스보고서에서 취업자의 정보를 이용하여 산출한다. 벤치마크수준은 지난벤치마크 기준해의 4월부터 현재벤치마크해의 3월까지의 각 월들의 표본조사기반 추정치들을 대체한다. 2002년 벤치마크해에는 표준산업분류(SIC : standard industrial classification)를 북미산업분류체계(NAICS : north american industry classification system)으로 전환하였으며 벤치마크수준들은 2001년1월부터 2002

년3월까지 추정치들을 대체하였다.

당해년 벤치마크수준을 위해서 신규 3월의 수준을 만드는데 적당한 표본연결함수를 신규수준에 적용한 후에 4월이후의 추정치들은 새로 계산되었다. 표본연결함수는 해당 월의 조사응답자들로부터 보고된 취업자를 전월의 동일조사응답자로부터 보고된 취업자로 나누어서 계산하여 표본추정치들의 월간변동을 추정한다. 벤치마크과정중에 사용된 연결함수는 고유 추정치를 산출할 때 사용했던 것과는 약간 다를 수가 있다. 왜냐하면 이전에 공표된 추정치에 포함되기에는 너무 늦게 보고된 응답자들의 자료도 포함되었기 때문이다. 지난 5년동안 비농업의 취업자의 총계에 대한 주단위 추정치의 벤치마크 수정값의 절대평균은 0.5%에서 0.9%사이 이다.

(3) 소영역 모형과 해당 오차의 측도

소영역 모형은 3개의 상이한 상대월간변동 추정치들($\widehat{L}_1, \widehat{L}_2, \widehat{L}_3$)의 가중합으로 구성된다. 3개의 상대월간 추정치들은 각 추정치들의 분산을 기반으로 한 가중계산한 것이다. 추정치의 분산이 클수록 그 추정치에 주어지는 가중값은 작아진다. 결론적으로 당월 취업자에 대한 추정치는 아래와 같이 정의한다.

$$\widehat{Y}_{iat} = [W_{iat,1} \widehat{L}_{iat,1} + W_{iat,2} \widehat{L}_{iat,2} + W_{iat,3} \widehat{L}_{iat,3}] \widehat{L}_{iat-1}$$

여기서 \widehat{L}_{iat} = a 소지역의 i 산업으로 정의 소영역의 t월의 취업자 추정치

\widehat{L}_{iat-1} = a 소지역의 i 산업으로 정의된 소영역의 t월의 전월 취업자 추정치

$\widehat{L}_{iat,1}$ = 소영역 ia에서 표본응답자를 기반으로 산출한 당월의 상대월간 변동 추정치

$\widehat{L}_{iat,2}$ = 소영역 ia에서 취업자 총계의 시계열적 특성을 이용한 예측값을 기반으로 산출한 당월의 상대월간 변동 추정치

$\widehat{L}_{iat,3}$ = 산업 i에서 소지역a를 포함하는 주에서 모든 표본응답자를 이용한 변동의 합성 추정치를 기반으로 산출한 당월의 상대월간 변동 추정치

$W_{iat,1}, W_{iat,2}, W_{iat,3}$ 은 3종류의 추정치 $\widehat{L}_{iat,1}, \widehat{L}_{iat,2}, \widehat{L}_{iat,3}$ 각각의 분산에 의해서 정해진 가중값들이다.

위의 추정모형에서 추정치 $\widehat{L}_{iat,1}, \widehat{L}_{iat,2}, \widehat{L}_{iat,3}$ 들중에서 하나 또는 두 개의 가중값이 0으로 지정될 가능성도 있다. 모형에 투입되는 추정치의 안정성 때문에 추정치에 0의 가중값을 부여하게 되는데 예를 든다면 $\widehat{L}_{iat,1}$ 또는 $\widehat{L}_{iat,3}$ 가 5명이하의 응답자를 기반으로 산출되었다면 이들에게 가중값으로 0을 부여하게 된다. 만일에 $\widehat{L}_{iat,2}$ 가 불안정한 분산을 갖거나 모형 적합도가 낮을 경우에도 0의 가중값을 부여하게 된다. 이와 같은 경우에는 소영역 모형의 추정치는 $\widehat{L}_{iat,1}, \widehat{L}_{iat,2}, \widehat{L}_{iat,3}$ 중에서 단지 하나 또는 두 개의 추정치들만의 기반에서 산출된다.

표본조사된 취업자들로부터 직접적으로 계산한 추정치에 대한 표본오차공식을 소영역 모형을 기반으로 산출한 추정치들에는 적용할 수 없다. 소영역 모형의 추정치에 대한 신뢰성을 평가할 수 있는 측도는 과거에 소영역 모형에 의한 추정값과 같은 기간중 취업자와 실업자의 총계에 대한 비교분석을 통해서 얻어질 수 있다. 이와 같은 측도는 유용하기는 하지만 과거의 소영역 모형에서 신뢰성이 좋았다고 해서 현재의 소영역 모형에서 신뢰성이 좋다는 보장은 없다.

2.5.2 OES데이터에 대한 소지역 추정법 적용사례

(1) 직업별 취업자 통계조사(occupational employment statistics survey)

미국 노동통계국의 직업별 취업자 통계조사의 프로그램은 지역과 산업분류 수준에서 특정직업에 대한 임금과 취업자의 추정치들을 계산하기 위해서 연간 우편조사로 설계되었다. 지역단위 추정치는 주단위와 대도시 통계단위(MSA:metropolitan statistical area)수준에서 이용 가능하고 임금과 직업별 취업자의 추정치는 국가단위에서 400개 산업에 대해서 이용 가능하다. OES표본들은 매년 400,000개 사업체에 대해서 농장근로자가 아닌 근로자들의 임금과 월급에 대한 자료를 700개 직업에 대해서 수집한다.

OES프로그램은 연방과 주정부의 협동프로그램으로 연방정부와 지방정부에서 정책결정에 이용하는 핵심적인 정보를 제공한다. 부가적으로 많은 관심지역단위들은 카운티 수준이다.

직업별 취업자(OE)와 임금의 추정치는 주단위, MSA와 3자리 산업분류(SIC)수준에서 정의된 기초단위까지 이용 가능하며 기초단위들의 합계수준에서도 이용 가능하다. 기초추정단위에서 가중표본 취업자는 미국 노동통계국의 종단적 데이터베이스로부터 추론한 전

체 취업자에 맞도록 조정한다. MSA의 하위단위와 같은 기초단위이하 수준의 추정치들은 조정되지 않기 때문에 편향이 있을 수 있으며 표본변동도 상당히 클 것이다. 지금부터 MSA의 하위단위 추정치를 통계적으로 좀더 정확하게 추정할 수 있는 3종류의 합성 추정량들을 살펴보자.

(2) 지역별 직업별 취업자에 대한 합성 추정량

① 기호

본 연구에서 이용될 기호를 아래에서 설명하겠다.

- E : 총 취업자. 예를 들면 $E_{d,g}$ 는 소영역 d 내에서 모집단 그룹 g 의 총 취업자이고, E_d 는 소영역 d 에서 총 취업자 수이다.
- $P_{..o}$: 직업 o 에 대한 OE의 퍼센트이고, $P_{d,g,o}$ 는 소영역 d 와 사업체 그룹 g 내에서 직업 o 에 대한 OE의 퍼센트이다. 여기서 $P_{..o}$ 는 어떤 수준에서도 알려져 있지 않기 때문에 비추정량을 이용하여 표본자료에서 추정해야한다.
- X_o : 다양한 수준에서 총 직업별 취업자이고 $X_{i,o}$ 는 사업체 I 에서 직업 o 의 총 OE이며 $X_{d,o}$ 는 소영역 d 에서 총 OE이다. 단지 몇몇 $X_{i,o}$ 만이 조사자료에서 이용 가능하므로 MSA 하위영역이나 소영역 수준에서 총 OE인 $X_{d,o}$ 의 추정을 살펴보자.
- w_i : 사업체 i 의 표본가중치로 무응답을 보정한 값이다.
- d : 추정영역, 소영역 또는 MSA의 하위영역을 나타내는 지표이다. ($d=1, \dots, D$)
- g : 사업체의 그룹을 나타내는 지표이며 일반적으로 지리적으로 그룹핑한 사업체 그룹을 말하지 않고 특별히 주내에서 동일한 2자리 산업분류에 속하는 모든 사업체들을 g 그룹으로 나타낸다. ($g=1, \dots, G$)
- h : MSA/SIC3 산업/규모로 정의된 사업체 그룹을 나타내는 지표이며, g 로 그룹핑한 것 보다는 더 세분화된 영역을 의미한다.
- i : 사업체의 지표이다.
- o : 직업을 나타낸다.
- $\{d\}$: 소영역 d 에 속하는 사업체의 그룹을 나타낸다.
- $\{g\}$: SIC2 산업분류 g 에 속하는 사업체의 그룹을 나타낸다.
- $\{h\}$: MSA/SIC3 산업분류/규모의 셀 h 에 속하는 사업체의 그룹을 나타낸다.

총 취업자, OE 또는 적합한 합계 수준에서 OE의 퍼센트의 추정치를 계산하는데 호르

비즈-툼슨 추정량을 이용한다.

② 합성 추정법

OES에서 지역 OE의 합성 추정량은 상업체의 유한 모집단에 대해서 고정된 표본설계를 가정한다. 현행 OES프로그램은 MSA/SIC3 산업분류/규모 수준에서 OE추정치를 제공한다. MSA의 하위영역인 d로 표시된 소영역에서 추정치를 계산하기 위해서는 그룹h내에서 총 취업자중에서 직업 o를 갖는 OE의 비율이 모든 소영역에서 일치한다는 추가적인 가정조건이 필요하다. 따라서 소영역 d에서 총 OE인 $X_{d,o}$ 를 아래와 같이 추정할 수 있다.

먼저 추정량 I 이라고 부를 첫 번째 합성추정량을 살펴보자.

$$\hat{X}_{d,o}^{(1)} = \sum_{h=1}^H E_{d,h} \cdot \hat{P}_{..h,o} \quad (1)$$

$$\text{여기서 } \hat{P}_{..h,o} = \frac{\sum_i w_i X_{i,o}}{\hat{E}_{..g}} \text{ 이고 } \hat{E}_{..g} = \sum_{o=1}^O \sum_i w_i X_{i,o} \text{이다.}$$

추정량 I 은 동일 그룹내에서 동질성의 가정이 충족될 경우에는 바람직한 추정치를 제공한다. 만일에 전제조건이 맞지 않으면 추정량II와 함께 편향된다. 이런 그룹은 상대적으로 세분되게 정의하므로 산업그룹이 희박한 카운티에 있는 사업체는 표본에 포함될 가능성이 적게 될 수 있어서 어떤 카운티에서는 추정치를 계산할 수 없을 것이다. 일반적으로 이런 너무 세분화된 그룹핑의 합성추정량은 효율이 낮게 된다. 과거의 경험적인 분석에서 OES의 세분화 그룹핑은 개선된 점을 보여주지 못하였다.

추정량 II은 주단위와 SIC2 산업분류내의 사업체들을 그룹핑 하였다. 여기서는 동일한 SIC 2 산업분류에 속한 사업체들은 동일한 OE분포를 갖는다고 가정하였다. SIC2 산업분류에서는 전체 취업자중 특정 OE의 비는 모든 지리적인 영역인 소영역에서 일치한다고 가정했음을 의미한다. 모든 g에서 $P_{..g,o} = P_{d,g,o}$ 가 성립한다. 이런 경우에는 추정량 I 과 유사하게 비 추정량을 이용하여 $X_{d,o}$ 에 대한 추정치를 계산할 수 있다.

$$\hat{X}_{d,o}^{(2)} = \sum_{g=1}^G E_{d,g} \cdot \hat{P}_{..g,o} \quad (2)$$

$$\text{여기서 } \hat{P}_{..g,o} = \frac{\sum_i w_i X_{i,o}}{\hat{E}_{..g}} \text{ 이고 } \hat{E}_{..g} = \sum_{o=1}^O \sum_i w_i X_{i,o} \text{이다.}$$

만일에 동질성에 대한 가설이 틀리게 된다면 추정량 II인 $\hat{X}_{d,o}^{(2)}$ 는 편향 추정량이다.

OE의 총계는 $X_{d,o} = \sum_{g=1}^G E_{d,g} \cdot P_{d,g,o}$ 이며 $\hat{X}_{d,o}^{(2)}$ 의 편향은 아래와 같이 나타낼 수 있다.

$$\begin{aligned}
bias &= X_{d,o} - E(\widehat{X}_{d,o}^{(2)}) \\
&= \sum_{g=1}^G E_{d,g} (P_{d,g,o} - P_{\cdot,g,o}) \quad (3)
\end{aligned}$$

만일 모든 사업체 그룹 g 에 대해서 $P_{d,g,o} = P_{\cdot,g,o}$ 가 성립하면 편향은 0이다. 실제로 있어서 이런 경우는 희박할 것이다.

만일 어떤 방법으로든지 편향을 계산할 수 있다면 편향을 교정한 합성추정량을 제안할 수 있을 것이다. 따라서 합성추정량 III 을 아래와 같이 쓸 수 있다.

$$\begin{aligned}
\widehat{X}_{d,o}^{(3)} &= \widehat{X}_{d,o}^{(2)} + \widehat{bias} \\
&= \widehat{X}_{d,o}^{(2)} + \sum_{g=1}^G \widehat{E}_{d,g} (\widehat{P}_{d,g,o} - \widehat{P}_{\cdot,g,o})
\end{aligned}$$

이식을 좀더 정리하고 추정치를 비 추정량으로 나타내면 추정량 III 은 아래와 같이 표현할 수 있다.

$$\widehat{X}_{d,o}^{(3)} = E_{d,\cdot} \left[\widehat{P}_{d,\cdot,o} + \sum_{g=1}^G \frac{E_{d,g}}{E_{d,\cdot}} - \frac{\widehat{E}_{d,g}}{\widehat{E}_{d,\cdot}} \right) \widehat{P}_{\cdot,g,o} \right] \quad (4)$$

이 형식에서는 소영역 d 에서 특정 OE의 추정은 소영역 d 에서 총 취업자의 비율로 나타내고 있으며 괄호 내에서는 소영역 d 에서 OE의 퍼센트의 추정치를 포함하고 있다.

(3) 시뮬레이션 연구 결과

세 추정량의 성능을 평가하고자 할 때 미국 내의 모든 사업체에 대한 완벽한 정보를 이용할 수 없으므로 시뮬레이션 연구가 필수적이다. 실제 모집단과 매우 유사한 직업별 취업자의 특성을 갖는 인위적인 사업체의 모집단을 구성하여 시뮬레이션으로 표본을 추출한다. 랜덤표본을 근거로 3개의 추정법에 대한 추정치를 계산하여 인위적인 모집단 모수의 참값과 비교 분석하였다.

인위적인 모집단은 미국의 대규모 주의 사업체들을 기반으로 생성하였다. 고려한 사업체들의 특성은 전체 사업체 취업자들의 분포, 전체 취업자에 대한 OE의 비율, MSA의 하위영역에 대한 OE의 분포 등이며 미국의 대규모주의 표본에 맞추어 모형화하였다. 시뮬레이션한 모집단에는 1000개의 사업체들은 갖는 두 개의 MSA안에는 6개 카운티를 포함시켰다. 50개 업종을 갖는 취업유형을 포함하고 사업체의 규모는 근로자 1명에서 3000명까지 7개 범주로 나누었다. 인위적인 모집단과 주의 표본사업체의 취업자 분포를 비교한 결과 노동력의 특성 항목에서 두 집단이 매우 유사한 것으로 나타났다. 1000개의 사업체를 3자리 산업분류코드와 규모 코드를 기준으로 총화하였으며 근로자수가 250명 이상인 사업체는 무조건 표본으로 추출하고 2명에서 250명 미만은 3자리 산업분류내에서 규

모층내의 근로자수에 비례하게 추출하였다. SIC3/규모층에 종사자가 한 명인 사업체만 포함되었을 경우에는 종사자가 한 명인 사업체도 표본에 포함시켰으며 각 SIC3/규모 층에서 사업체들은 계통추출법으로 선정하였다.

표본을 추출할 때마다 3종류의 추정량에 따라서 각 카운티 내의 직업별로 추정치를 계산하였으며 각 표본마다 6개 카운티에서 263개 추정치가 계산되었다. 각 표본은 250개 사업체를 포함하였으며 전체적으로 200개 독립표본을 추출하여 모든 추정치들은 평균을 계산하였고 결과는 <표 1>에 주어졌다. <표 2>에는 한 카운티에서 특정직업에 대한 추정치를 요약하였다.

< 표 2 > 6개 카운티에서 OE 추정치 비교

구분	추정량 I	추정량 II	추정량III
평균 표준오차	3.049	12.172	2.508
평균 변동계수	0.090	0.096	0.087
평균 편향	-2.54	-2.534	-0.104
평균 상대편향	-0.027	-0.087	0.051
전체 오차율	0.059	0.137	0.05

<표 1>에서 추정량 II의 평균표준오차가 추정량 I 과 추정량III에 비해서 매우 크게 나타났으며 유사한 사업체들을 크게 그룹핑하여 변동이 커진 것으로 보인다.

< 표 3 > 한 카운티에서 특정 직업의 추정치

구분	추정량 I	추정량 II	추정량III
95% C.I. nominal coverage	0.942	0.955	0.968
표준화오차 신뢰경계추정	(-0.513, 0.39)	(-0.32, 0.14)	(-0.45, 0.47)
bias confidence bound estimate	(-39.2, 29.9)	(-48.6, 20.8)	(-34.9, 36.1)

추정량III은 다른 추정량들에 비해서 편향이 반 정도 크기인 것은 편향을 교정한 효과로 생각된다. 그러나 3 추정량의 평균 변동계수는 유사한 경향을 보인다. 전체적인 오차율(OER)은 아래와 같이 정의하였으며 추정량II가 유의할 정도로 크게 나타났다.

$$OER = \frac{\|E(\hat{X}) - X\|}{\|X\|}$$

OER은 벡터 추정치를 참값과 비교한 퍼센트오차의 측도이며 3개 추정량이 강력한 신뢰구간범위를 보이고 있다.

또 다른 흥미로운 특징은 표본추출률이 커지면 추정량의 오차율이 감소할 것으로 기대하였으나 추정량Ⅱ는 다른 추정량들과 비교하여 감소율이 느리게 나타났다. 전체적으로 표본설계와 추정법의 효과는 복잡하게 나타나고, 추정량Ⅰ과 추정량Ⅲ에서 소규모 셀의 영향이 컸으며 빈 셀을 처리할 수 있는 프로그램에 유의해야 한다는 것이다.

(4) 결론

3개의 합성 추정량을 비교한 결과에서 신뢰구간범위와 변동계수에서는 유사한 특성을 보였으며 추정량Ⅲ은 좀더 적은 편향을 나타냈고 추정량Ⅱ는 가장 큰 평균표준오차를 갖는다. 진짜모집단과 유사한 인위적인 모집단의 시뮬레이션에 대한 연구가 더 필요하며 소지역 모형을 결합하여 3가지 합성 추정량을 실제 자료에 적용하는 연구는 좀 더 심층적으로 추진할 만한 가치가 있을 것이다.

제 3 장 현행 임금구조기본통계조사 표본 분석

3.1 현행 임금구조통계조사 표본설계 개요

현행 임금구조기본통계조사를 위한 표본설계는 2002년 6월에 새로 이루어졌다. 이 표본설계에서는 산업중분류를 층화변수로 사용하였고, 각 산업중분류 구분에서 사업체 규모에 따라 2차로 층화하였다. 따라서 60개 산업중분류와 6개의 사업체 규모 구분을 고려하면 모두 360개 층으로 구성되었다. 또한 사업체를 1차 추출단위로 하고, 표본 사업체 내의 상용근로자를 2차 추출단위로 하는 층화 2단계추출법에 의해서 추출되었다. 사업체에 대한 추출틀은 2000년 12월 말 기준으로서 조사된 사업체노동실태조사 결과를 기초로 구성되었다.

표본추출에 있어서 각 층에 대한 표본 사업체는 각 산업중분류별 사업체 규모에 배정된 표본크기만큼을 계통추출하였다. 한편 사업체 추출과정에서 표본의 지역 대표성 확보를 위해 지방노동사무소, 행정구역에 따라 사업체를 정렬한 후 계통추출법을 적용하였다. 한편 조사 근로자는 선정된 표본 사업체의 상용근로자 명부에서 각 사업체 규모에 따라서 주어진 조사 근로자 추출률에 의거하여 계통추출하였다. 이에 따른 표본설계시에 표본으로 추출된 사업체 수는 6,344개이고 예상 조사 근로자수는 424,542명이었다.

임금구조통계조사 표본설계 당시의 산업중분류별 표본 사업체 수, 사업체 규모별 표본 사업체 수 등 표본설계와 관련된 자세한 사항은 2002년 6월에 발간된 “임금구조통계조사 표본설계 최종보고서”를 참고하기 바란다.

3.2 현행 임금구조통계조사 자료 분석

현행 표본설계에 따라 구성된 표본을 대상으로 2002년 8월부터 9월까지 1.5개월에 걸쳐 조사된 자료를 기초로 분석한 결과 조사 완료된 사업체 수는 5,988개이고, 조사 근로자수는 380,292인 것으로 나타났다. 실제 2002년도 임금구조통계조사에서 조사완료된 사업체 및 근로자를 산업분류 및 직종분류에 따라 구분하여 정리하고, 아울러 산업분류 또는 직종분류별 근로자 평균임금을 산출한 결과는 다음과 같다.

본 연구에서 평균임금은 각 근로자의 임금총액을 기준으로 표본설계 가중값을 적용하여 구한 가중평균임금이다. 임금총액에는 정액급여, 초과급여 및 연간특별급여가 포함된 것이다. 즉 조사 자료에서 임금총액은 다음과 같이 산출된 것이다.

$$\cdot \text{임금총액} = \text{정액급여} + \text{초과급여} + \text{연간특별급여}/12$$

(1) 산업분류별 표본 현황

▶ 산업분류별 표본 사업체 수, 근로자 수 및 평균임금

산업 중분류 및 소분류별 조사완료 사업체 수, 근로자 수 및 월 평균임금은 <표 3.1>과 같다. <표 3.1>을 보면 농업 및 임업의 경우 중분류 01(농업)의 경우 조사완료 근로자 수는 2,033명이고, 해당 표본 사업체 수는 81개인 것으로 나타났으며 이들 2,033명의 월 평균임금총액은 1,865,957원이었다. 한편 중분류 01(농업)을 세분하여 산업소분류별로 정리해 보면 소분류 011(작물재배업), 012(축산업) 및 014(조경수 식재 및 농업관련 서비스업)에 속하는 표본 근로자 수는 각각 344명, 536명, 1,153명이었고, 표본 사업체 수는 각각 16개, 29개, 36개인 것으로 나타났으며, 각 산업소분류별 월평균임금총액은 1,722,686원, 1,329,650원, 2,190,468원인 것으로 파악되었다.

▶ 산업소분류별 평균임금 산출을 위한 소지역 추정기법의 필요성

<표 3.1>을 보면 예를 들어 농업의 산업소분류 중 013(작물재배 및 축산 복합 농업) 및 015(수렵 및 관련 서비스업) 등의 경우 현행 표본에서 해당 사업체에 대한 조사가 전혀 이루어지지 않고 있기 때문에 현행 표본조사 자료를 직접 사용하여 해당 산업소분류에 대한 평균임금 산출은 불가능하다. 아울러 소분류 111, 112, 231, 233, 294, 403 등의 경우를 보면 표본 사업체 수가 1-3개로 매우 적기 때문에 이 자료를 기초로 한 해당 산업소분류에 대한 평균임금 추정값은 신뢰성에 있어서 상당한 문제가 발생하게 된다. 따라서 현행 표본조사를 통해 소수의 사업체만이 조사되고 있거나 혹은 전혀 사업체가 조사되지 못하고 있는 산업소분류에 대한 평균임금통계의 산출을 위해서는 적절한 소지역 추정기법의 활용이 필요하다.

한편 적절한 소지역 추정기법의 활용하는 경우에도 표본조사에서 얻어지는 자료가 전혀 없어 직접 추정값을 산출할 수 없는 경우 보조정보가 완벽하지 않으면 추정값의 신뢰성에 문제가 발생한 소지가 있다. 따라서 산업소분류별 평균임금통계 산출을 위한 소지역 추정기법의 적용에 있어서도 모든 산업소분류의 평균임금 산출을 목표로 하는 것보다는 현실적으로 관계자 협의를 통해 일부 중요하지 않은 산업소분류는 추정 대상에서 제외하는 방안을 고려하는 것이 필요할 것으로 판단된다.

<표 3.1> 산업분류별 표본 근로자수, 사업체수 현황 및 평균임금

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금				
농업 및 임업	1	2,033	81	1,865,957	11	344	16	1,722,686				
					12	536	29	1,329,650				
					14	1,153	36	2,190,468				
어업	2	464	29	2,179,151	20	464	29	2,179,151				
					5	466	27	1,667,091				
광업	5	466	27	1,667,091	51	235	12	2,219,749				
					52	231	15	1,253,570				
제조업	10	3,463	72	1,935,094	101	3,463	72	1,935,094				
					11	72	4	1,958,646				
					111	63	3	2,058,469				
					112	9	1	1,237,703				
제조업	12	2,339	105	1,864,859	121	2,214	100	1,881,261				
					122	125	5	1,518,747				
	15	14,132	200	1,736,155	151	4,205	72	1,351,370				
					152	1,081	8	2,058,807				
					153	985	30	2,102,012				
					154	5,829	72	1,724,251				
					155	2,032	18	2,252,257				
					16	1,647	9	3,346,986				
					17	9,902	152	1,471,344	160	1,647	9	3,346,986
									171	5,223	44	1,563,533
	172	1,508	37	1,413,727								
	173	339	17	1,470,850								
	174	2,405	34	1,468,085								
	179	427	20	1,366,590								
	18	7,901	140	1,243,334					181	7,901	140	1,243,334
									19	3,756	54	1,352,986
					191	1,325	15	1,474,362				
					192	982	16	1,531,397				
	20	2,988	50	1,582,966	193	1,449	23	1,097,962				
					201	393	16	1,548,665				
					202	2,595	34	1,600,024				
	21	4,428	62	1,844,731	211	2,784	27	1,979,292				
					212	1,644	35	1,711,211				
	22	9,007	176	1,930,936	221	6,155	84	2,097,093				
222					2,534	87	1,701,427					
223					318	5	2,063,719					
23	4,228	45	3,054,781	231	37	3	1,838,346					
				232	3,917	41	3,077,813					
				233	274	1	3,049,543					
24	11,980	105	2,365,373	241	3,891	48	2,599,119					
				242	3,235	19	2,224,997					
				243	2,894	32	1,884,409					
				244	1,960	6	3,128,951					

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금
	25	7,690	88	1,716,050	251	3,204	19	1,871,988
					252	4,486	69	1,651,426
	26	9,651	104	1,854,270	261	3,806	17	2,214,380
					262	2,062	20	1,663,122
					263	3,055	42	1,910,016
					269	728	25	1,490,819
	27	6,672	59	1,917,563	271	4,921	36	1,995,747
					272	1,514	14	1,836,661
					273	237	9	1,530,406
	28	5,676	75	1,652,130	281	1,110	21	1,711,059
					289	4,566	54	1,630,616
	29	10,122	119	1,837,769	291	3,671	52	1,732,937
					292	1,113	10	1,936,578
					293	2,148	45	1,780,048
					294	1,069	2	3,015,170
					295	2,121	10	1,956,561
	30	4,670	49	1,922,495	300	4,670	49	1,922,495
	31	9,403	116	1,542,082	311	3,200	31	1,581,087
					312	2,610	37	1,454,200
					313	1,640	12	1,779,050
					314	598	9	1,723,205
					315	295	13	1,277,983
					319	1,060	14	1,550,033
	32	12,094	94	1,849,678	321	6,732	50	1,847,879
					322	2,078	23	1,918,487
					323	3,284	21	1,791,318
	33	4,702	84	1,648,640	331	626	19	1,461,665
					332	1,897	38	1,852,590
					333	1,583	21	1,450,260
					334	596	6	1,542,327
	34	9,984	78	2,107,311	341	899	5	2,223,926
					343	9,085	73	2,096,661
	35	13,900	67	2,296,932	351	9,346	46	2,431,129
					352	1,347	9	1,934,451
					353	2,532	6	2,269,283
					359	675	6	1,602,120
	36	4,462	53	1,387,578	361	2,360	26	1,423,730
					369	2,102	27	1,359,027

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금
	37	763	48	1,516,912	371	435	27	1,557,359
					372	328	21	1,471,421
	40	5,284	49	3,128,230	401	4,217	32	3,235,552
					402	918	15	2,665,422
전기,가스 및 수도사업	41	1,394	29	3,095,074	403	149	2	3,124,609
					410	1,394	29	3,095,074
건설업	45	8,672	119	1,996,592	451	4,013	72	1,843,717
					452	4,659	47	2,192,434
	46	5,315	180	1,672,573	461	1,990	70	1,579,907
					462	519	31	1,593,704
					463	2,118	57	1,776,546
					464	666	19	1,842,060
					465	22	3	1,196,922
도매 및 소매업	50	2,696	147	1,754,904	501	1,192	65	2,143,221
					502	921	33	1,668,215
					503	46	3	1,913,215
					504	537	46	1,238,953
	51	10,157	232	2,008,664	511	1,151	33	2,454,652
					512	175	7	1,723,209
					513	1,534	42	1,938,054
					514	2,380	62	1,707,198
					515	637	18	2,177,728
	52	9,823	150	1,601,072	516	456	7	2,239,630
					517	458	14	1,887,911
					518	2,793	41	2,098,383
					519	573	8	2,408,186
					521	7,490	53	1,704,224
					522	140	10	1,269,149
					523	206	17	1,511,742
					524	371	13	1,585,718
					525	192	16	1,496,402
					526	493	34	1,526,678
숙박 및 음식점업 운수업	55	9,184	133	1,250,289	528	931	7	1,720,150
					551	6,878	46	1,424,600
	60	11,011	100	1,543,071	552	2,306	87	1,152,014
					602	10,363	75	1,505,503
					603	648	25	1,760,517
					611	2,411	45	2,272,763
					612	215	3	1,730,954

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금				
	62	2,556	46	2,860,801	621	2,218	35	2,910,132				
					622	338	11	2,116,736				
	63	10,740	187	1,810,229	631	1,769	18	2,199,439				
					632	869	34	1,901,724				
					633	1,163	45	1,316,047				
					639	6,939	90	1,863,766				
통신업	64	5,584	42	3,186,566	641	22	3	2,292,560				
					642	5,562	39	3,206,896				
금융 및 보험업	65	7,680	165	2,666,706	651	5,155	104	2,891,533				
					659	2,525	61	2,228,982				
	66	3,643	45	2,588,792	660	3,643	45	2,588,792				
					671	4,294	37	3,888,833				
					672	268	9	2,773,302				
부동산 및 임대업	70	7,716	231	1,208,866	701	2,319	37	1,883,863				
					702	5,397	194	1,071,559				
	71	1,571	37	1,946,778	711	148	8	1,151,960				
					712	1,317	24	2,075,898				
					713	106	5	2,351,783				
사업 서비스업	72	10,635	160	2,247,987	721	1,256	8	2,286,934				
					722	8,265	139	2,257,197				
					723	432	4	2,593,048				
					724	682	9	1,948,916				
	73	7,039	48	3,176,814	731	6,831	41	3,204,372				
					732	208	7	2,912,864				
					74	14,125	386	2,000,793	741	2,348	147	1,767,293
									742	1,549	29	2,349,841
									743	7,034	125	2,209,638
	75	19,541	236	1,177,060	744	1,738	57	1,893,396				
					745	699	14	2,154,482				
					746	328	4	1,689,163				
					749	429	10	1,100,678				
751					9,502	103	1,161,089					
교육 서비스업	80	12,278	194	2,154,322	759	10,039	133	1,192,808				
					801	720	55	1,280,203				
					802	1,770	30	2,445,546				
					803	8,180	38	3,009,316				
					804	204	4	2,250,071				
809	1,404	67	1,441,819									

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금
보건 및 사회복지사업	85	7,600	78	1,863,846	851	7,600	78	1,863,846
	86	1,248	55	1,250,897	861	632	20	1,268,275
					862	616	35	1,241,298
오락,문화 및 운동관련 서비스업	87	3,174	47	2,446,154	871	1,343	23	2,429,052
					872	1,720	21	2,541,113
					873	111	3	1,334,333
	88	4,611	48	1,685,664	881	326	1	3,463,795
					882	208	3	1,211,230
					883	1,919	32	1,632,848
					889	2,158	12	1,791,097
					889	2,158	12	1,791,097
기타 공공 수리 및 개인 서비스업	90	2,097	48	1,690,007	901	150	10	1,529,965
					902	1,947	38	1,714,544
	91	4,114	138	1,917,099	911	1,844	58	2,180,527
					912	97	4	1,990,743
					919	2,173	76	1,719,845
					919	2,173	76	1,719,845
	92	4,036	167	1,505,721	921	1,012	43	1,739,753
					922	2,809	105	1,408,697
					923	215	19	1,540,310
					923	215	19	1,540,310
93	985	52	1,184,814	931	400	24	1,108,642	
				939	585	28	1,243,642	

(2) 직종분류별 표본 현황

▶ 직종분류별 표본 사업체 수, 근로자 수 및 평균임금

현행 표본조사자료에서 직종 중분류 및 소분류별 조사완료 사업체 수, 근로자 수 및 월 평균임금은 <표 3.2>와 같다. <표 3.2>를 보면 예를 들어 전문가 중 중분류 11(과학전문가)의 경우 조사완료 근로자수는 958명이고, 해당 표본 사업체 수는 58개인 것으로 나타났다. 이들 958명의 월 평균임금총액은 3,273,904원이었다. 한편 중분류 11(전문가)을 세분하여 직종소분류별로 정리해 보면 소분류 111(자연과학 전문가), 112(생명과학 전문가) 및 113(사회과학 전문가)에 속하는 표본 근로자 수는 각각 564명, 317명, 77명이었고, 표본 사업체 수는 각각 17개, 29개, 12개인 것으로 나타났으며, 각 직종소분류별 월평균임금총액은 4,153,872원, 2,664,606원, 3,171,346원인 것으로 파악되었다.

▶ 직종세분류별 평균임금 산출을 위한 소지역 추정기법의 도입

현행 임금구조통계조사에서는 직종분류에 있어서 직종소분류까지 조사하고 있기 때문에 현행 조사자료는 본 연구에서 목표로 하는 직종세분류별 평균임금 산출을 위한 소지역 추정기법의 개발에는 한계가 있다. 따라서 직종세분류별 평균임금 산출을 위한 통계기법개발을 위해서 우선 2003년 임금구조통계조사에서 실험적으로 일부 산업에 대해 실험적으로 직종세분류를 조사하고 이 자료를 토대로 차후 좀더 심층적인 소지역 추정기법에 대한 연구가 이루어질 계획이다.

한편 직종세분류별 평균임금 산출에 있어서도 일부 세분류의 경우 실제 표본조사에서 얻어지는 자료가 전혀 없어 직접 추정값을 산출할 수 없는 경우 보조정보가 완벽하지 않으면 소지역 추정기법의 신뢰성에 문제가 발생한 소지가 있다. 따라서 직종세분류별 평균임금통계 산출을 위한 소지역 추정기법의 적용에 있어서도 실제 관심대상이 되는 주요 직종세분류를 선정하여 이들 선정된 직종세분류에 대해서만 제한적으로 평균임금 산출을 목표로 하는 방안을 고려해 보는 것이 필요할 것으로 판단된다.

<표 3.2> 직종분류별 표본 근로자 수, 사업체 수 현황 및 평균임금

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금
의회의원, 고위임직원 및 관리자	2	9,765	3,935	3,371,452	21	868	663	4,152,826
					23	3,259	1,510	3,106,742
					24	5,638	1,762	3,381,679
	3	3,411	2,705	2,967,129	30	3,411	2,705	2,967,129
전문가	11	958	58	3,273,904	111	564	17	4,153,872
					112	317	29	2,664,606
					113	77	12	3,171,346
	12	7,997	419	2,640,618	120	7,997	419	2,640,618
	13	15,428	1,112	2,693,015	131	5,939	283	2,426,388
					132	6,834	442	2,877,310
					133	1,097	179	2,875,589
					134	326	44	2,077,977
					135	1,232	164	3,346,494
					14	4,123	591	2,168,276
	15	6,926	222	2,344,731	142	183	74	1,917,150
					143	2,675	180	1,757,868
					144	191	56	1,594,559
					145	272	205	1,344,823
					151	4,361	39	3,690,447
					152	1,439	31	2,585,316
					153	85	5	2,588,023
	16	4,482	1,079	3,610,300	154	525	89	971,052
					155	99	4	2,297,595
					156	319	42	1,431,703
					157	98	12	3,406,820
					161	1,268	383	3,127,614
	17	387	90	2,049,775	162	1,152	196	3,355,598
					163	216	159	2,914,797
					164	940	70	4,487,206
					165	905	270	3,372,045
					167	1	1	6,205,000
					171	98	29	4,406,704
	18	4,969	423	2,137,528	172	210	37	1,399,623
					173	79	24	1,950,263
181					155	54	2,334,157	
182					2,468	108	2,424,477	
				183	1,476	224	1,655,367	
				184	870	37	3,034,406	

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금					
기술공 및 준전문가	21	848	110	2,428,038	211	406	31	2,948,193					
					212	333	67	2,202,414					
					213	109	12	2,063,812					
	22	6,284	779	1,959,440	220	6,284	779	1,959,440					
					23	32,735	3,242	2,079,624					
						232	11,863	900	2,406,821				
						233	2,452	292	2,339,525				
						234	2,351	318	1,739,853				
						235	1,334	139	1,921,661				
						236	661	48	2,921,097				
						237	6,339	940	1,965,700				
						238	574	110	2,182,826				
						24	1,344	179	1,344,769	241	1,191	138	1,284,453
										242	143	37	1,727,917
										243	10	4	1,597,141
	25	1,229	97	1,358,726	251	601	30	1,330,351					
					252	172	28	1,353,299					
					253	456	39	1,372,552					
	26	7,534	1,079	2,829,347	261	3,158	165	3,640,633					
					262	4,006	824	2,415,914					
					263	370	90	1,962,654					
	27	607	59	1,317,001	271	321	37	1,257,707					
					272	286	22	1,509,741					
	28	503	97	1,553,965	281	108	13	1,197,515					
					282	381	80	1,582,103					
					283	14	4	1,482,215					
	29	11,262	2,596	2,649,472	291	10,526	2,431	2,693,320					
					292	736	165	2,087,124					
	사무 종사자	31	76,631	13,084	1,818,109	311	19,639	2,619	2,004,319				
						312	6,945	1,056	2,251,652				
314						13,150	1,585	1,914,172					
315						15,273	4,052	1,679,644					
316						13,638	1,982	1,779,585					
317						7,735	1,680	1,410,204					
318						251	110	1,794,103					
32						8,415	1,219	1,651,986	321	3,627	527	1,683,674	
		322	2,616	387	1,416,319								
		323	2,172	305	1,809,179								

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금	
서비스 종사자	41	1,776	193	1,233,043	411	333	48	1,278,465	
					412	279	46	1,124,185	
					413	400	39	1,048,943	
					414	26	10	2,363,669	
					415	483	27	1,490,372	
					416	255	23	1,340,770	
	42	7,029	864	1,160,761	421	3,940	698	1,194,502	
					422	3,089	166	1,107,869	
	43	511	38	1,694,548	431	403	23	1,906,558	
					432	108	15	1,297,262	
	44	1,735	173	2,078,471	442	100	31	2,552,320	
					444	1,635	142	2,041,315	
	판매 종사자	51	6,567	483	1,533,484	511	2,099	166	1,691,675
						512	3,550	259	1,336,948
						513	918	58	2,014,732
		52	489	31	1,610,044	521	412	18	1,270,828
522						77	13	2,066,677	
53		92	9	1,218,013	530	92	9	1,218,013	
농업, 임업 및 어업숙련 종사자	61	864	145	1,652,168	611	110	9	2,316,566	
					612	6	2	1,566,627	
					613	301	86	1,569,863	
					614	18	4	1,764,380	
					615	415	40	1,532,346	
					616	13	3	861,166	
					617	1	1	965,625	
	62	33	11	1,564,612	620	33	11	1,564,612	
	63	111	19	1,492,435	630	111	19	1,492,435	
	기능원 및 관련기능 종사자	71	3,345	478	1,669,574	711	851	89	1,841,835
712						766	164	1,547,526	
713						924	142	1,625,659	
714						804	83	1,785,685	
72		6,292	408	1,796,592	721	5,022	325	1,816,245	
					722	1,270	83	1,711,073	
73		18,178	1,818	1,949,155	731	6,446	493	1,693,110	
					732	3,511	446	2,149,601	
					733	8,221	879	2,042,223	
74		4,233	287	1,515,364	741	1,500	60	1,308,910	
					742	1,787	82	1,689,490	
					743	121	27	1,679,983	
					744	825	118	1,625,491	

대분류	중분류	근로자수	사업체수	가중평균임금	소분류	근로자수	사업체수	가중평균임금
	75	10,516	452	1,217,502	751	2,524	114	1,260,179
					752	351	57	1,365,125
					753	6,399	243	1,205,033
					754	1,242	38	1,073,858
장치,기계 조작 및 조립 종사자	81	14,869	992	1,971,041	811	862	99	1,833,259
					812	2,800	109	1,823,707
					813	1,956	27	1,826,345
					814	1,688	62	1,683,035
					815	3,412	169	2,424,486
					816	2,988	480	2,031,178
					817	1,163	46	1,714,303
	82	27,083	1,180	1,639,699	821	5,689	354	1,701,364
					822	3,392	155	1,698,708
					823	3,261	123	1,532,017
					824	433	37	1,470,097
					825	2,637	125	1,832,913
					826	5,651	172	1,493,224
					827	2,849	85	1,805,332
					828	1,232	34	1,975,896
					829	1,939	95	1,447,389
					83	19,304	533	1,593,671
	832	11,558	270	1,549,562				
	833	1,180	77	1,306,059				
	84	17,965	2,009	1,593,316	841	306	17	2,303,824
842					13,716	1,506	1,539,012	
843					3,636	452	1,964,830	
844					307	34	2,089,323	
단순노무 종사자	91	19,340	1,981	954,790	911	5,888	554	892,703
					912	8,994	1,085	925,602
					913	1,137	122	1,387,896
					914	3,055	164	963,068
					915	266	56	1,270,265
	92	324	40	970,032	920	324	40	970,032
	93	11,492	769	1,172,897	930	11,492	769	1,172,897
	94	2,306	188	1,425,684	941	1,292	98	1,431,668
					942	1,014	90	1,422,680

(3) 산업분류 및 직종분류에 따른 표본 현황

근로자 평균임금은 동일 산업분류에 속하는 근로자의 경우에도 근로자의 직종에 따라 임금에 차이가 발생하게 된다. 효율적인 소지역 추정기법의 개발을 위해서는 동일 산업분류에 속하는 근로자들의 임금이 직종에 따른 유의적인 차이가 발생하는지 검토할 필요가 있으며, 이를 소지역 추정모형에 적절한 형태로 반영하는 것이 필요하다. 이런 분석작업은 실제 각 산업대분류 및 산업중분류별로 이루어져야 하고 이들 분석결과는 매우 방대하기 때문에 이들 결과를 보고서에 모두 수록하는 것은 현실적으로 적절치 않다. 따라서 <표 3.3>에는 대표적인 산업대분류에 해당하는 제조업의 경우로 한정하여 제조업의 각 산업중분류내에서 직종에 따른 근로자 수 및 해당 사업체 수와 평균임금을 정리하였다.

▶ 제조업 산업중분류별 직종분류에 따른 표본현황 및 평균임금

<표 3.3>을 보면 제조업중 산업중분류 15(음식료품 제조업)에 해당하는 표본 근로자수는 14,132명이고(<표 3.1> 참조) 그 중 직종 중분류에 있어서 02(행정 및 경영관리자)에 속하는 근로자수는 332명(96업체)이고 이들의 평균임금은 3,435,829원이다. 한편 이를 직종에 따라 세분하여 보면 직종중분류 02에 속하는 표본 근로자 332명 중 직종소분류 021(기업 고위임원), 023(생산 및 운영부서 관리자), 024(기타부서 관리자)에 속하는 표본 근로자수(사업체수)는 각각 29명(26개 업체), 122명(78개 업체), 181명(70개 업체)이고, 평균임금은 각각 4,452,548원, 3,143,244원, 3,441,396원인 것으로 나타났다.

▶ 산업소분류별 소지역 추정모형에서 직종분류의 활용

제시된 <표 3.3>을 보면 제조업에 대한 제한적인 결과이지만 동일 산업중분류에 속하는 근로자라고 하더라도 그 근로자가 어떤 직종에 속하는지에 따라 평균임금에 있어서 상당한 차이를 보여 주고 있다. 따라서 산업소분류별 평균임금 산출을 위한 소지역 모형을 개발하는 데 있어서 직종분류 변수를 활용함으로써 보다 효율적인 소지역 추정 모형의 구현이 가능해질 수 있다는 사실을 예상할 수 있다.

특히 각 산업소분류에 속하는 근로자들에 대한 직종분류별 구성비율을 다른 통계 또는 전수조사를 통해 보조정보로 확보할 수 있다면 매우 효율적인 소지역추정기법의 구현이 가능해질 수 있다. 하지만 현실적으로 이와 같은 정보를 제공해 줄 수 있는 통계는 존재하지 않기 때문에 본 연구에서는 임금구조통계조사에서 나타난 산업분류별 직종구성비율을 적절하게 활용하는 방안을 강구하고자 한다.

<표 3.3> 제조업 산업중분류내 직종중소분류별 표본 현황 및 평균임금

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
15	2	332	96	3,435,829	21	29	26	4,452,548
					23	122	78	3,143,244
					24	181	70	3,441,396
	3	100	83	2,693,636	30	100	83	2,693,636
	11	9	4	3,804,140	111	1	1	4,447,800
					112	8	3	3,688,131
	12	17	8	2,661,132	120	17	8	2,661,132
	13	139	22	3,198,361	131	1	1	3,924,317
					132	8	5	2,701,365
					133	63	14	2,273,622
					135	67	7	4,430,962
	14	63	33	1,319,788	141	4	4	2,194,309
					142	2	1	2,735,080
					143	11	10	1,534,212
					145	46	22	1,169,796
	16	87	28	3,311,516	161	44	18	2,988,224
					162	5	4	3,597,656
					163	10	9	3,229,253
					164	4	3	4,362,155
					165	24	10	3,711,441
	17	2	1	4,549,596	171	2	1	4,549,596
	18	7	6	1,457,551	183	7	6	1,457,551
	21	37	7	2,342,068	211	1	1	2,089,500
					212	36	6	2,353,481
	22	58	22	2,059,265	220	58	22	2,059,265
	23	414	70	2,194,721	231	7	4	1,865,402
					232	54	26	2,384,014
					233	50	13	2,022,834
					234	6	4	2,000,386
					235	4	1	3,377,704
					236	2	1	2,503,667
					237	259	55	2,126,632
					238	32	5	2,615,717
	24	6	6	1,998,160	241	4	4	1,657,025
					242	2	2	2,156,963
	25	1	1	2,415,000	253	1	1	2,415,000
	26	182	34	2,263,006	261	7	4	3,181,388
					262	166	31	2,234,667
					263	9	4	2,166,780

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	28	4	1	1,810,546	282	4	1	1,810,546
	29	333	102	2,553,528	291	325	101	2,560,250
					292	8	5	2,064,047
	31	2773	173	1,728,262	311	464	96	1,831,696
					312	154	35	1,978,948
					314	696	71	1,800,887
					315	455	144	1,386,112
					316	779	104	1,994,429
					317	225	56	1,203,706
	32	51	14	1,666,916	321	17	4	1,185,851
					322	5	5	1,658,076
					323	29	8	1,928,492
	41	2	2	1,752,886	411	2	2	1,752,886
	42	232	43	1,029,389	421	158	42	991,597
					422	74	2	1,283,441
	43	2	1	1,624,875	431	2	1	1,624,875
	44	1	1	2,705,283	442	1	1	2,705,283
	51	191	26	1,534,634	511	93	17	1,853,497
					512	98	10	1,176,217
	52	5	1	2,327,533	521	5	1	2,327,533
	53	9	1	630,583	530	9	1	630,583
	61	32	5	1,199,780	613	2	2	1,434,351
					615	22	2	1,666,746
					616	8	1	840,000
	71	36	13	2,209,548	711	1	1	783,750
					712	3	3	2,031,805
					713	29	9	2,371,137
					714	3	3	2,299,491
	72	12	5	1,955,750	721	12	5	1,955,750
	73	311	57	2,268,524	731	32	10	2,247,147
					732	137	36	2,259,245
					733	142	31	2,285,053
	74	3	2	2,194,717	744	3	2	2,194,717
75	2217	86	1,294,789	751	2217	86	1,294,789	
81	196	39	2,319,039	812	1	1	2,205,000	
				815	8	5	1,844,055	
				816	167	35	2,389,728	
				817	20	3	2,060,525	

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금				
	82	3931	99	1,819,630	821	21	5	2,256,787				
					822	72	2	1,303,604				
					823	2	2	1,898,882				
					824	3	2	3,409,295				
					825	23	2	1,580,914				
					827	2765	75	1,869,362				
					828	573	24	1,741,535				
					829	472	21	1,684,409				
	83	12	5	2,110,416	831	4	3	2,604,371				
					832	8	4	1,837,032				
	84	349	83	1,840,613	842	214	75	1,550,771				
					843	135	26	2,481,688				
	91	188	64	1,277,025	911	46	22	1,307,312				
					912	130	49	1,293,580				
					913	8	4	1,124,987				
					914	3	2	1,483,931				
					915	1	1	850,850				
	92	23	4	930,850	920	23	4	930,850				
	93	1754	85	1,243,771	930	1754	85	1,243,771				
	94	11	3	1,983,817	942	11	3	1,983,817				
17	2	168	64	3,067,093	21	13	11	4,165,954				
					23	80	51	2,881,556				
					24	75	39	3,011,067				
	3	69	63	2,742,239	30	69	63	2,742,239				
	12	13	4	2,588,876	120	13	4	2,588,876				
					13	62	15	2,815,957	132	6	4	2,383,130
									133	52	11	2,246,129
	135	4	3	4,144,074								
	14	8	7	1,343,873	143	4	4	1,586,409				
					145	4	4	1,201,473				
	15	3	2	1,230,500	154	3	2	1,230,500				
	16	23	11	3,145,938	161	7	6	3,014,321				
					162	3	3	3,592,454				
					163	5	4	3,021,777				
					165	8	5	3,057,522				
	18	13	5	1,153,730	182	1	1	621,000				
183					12	4	1,159,445					
22	10	5	1,944,420	220	10	5	1,944,420					

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	23	253	42	1,893,278	231	4	3	2,116,104
					232	12	5	2,190,792
					233	66	17	1,783,120
					234	19	6	1,139,094
					235	1	1	2,744,042
					237	137	26	1,890,256
					238	14	6	2,195,955
	24	3	3	1,788,784	241	3	3	1,393,332
	25	12	1	2,095,454	253	12	1	2,095,454
	26	112	19	2,188,065	262	111	19	2,173,600
					263	1	1	2,166,911
	27	3	1	1,469,310	271	3	1	1,469,310
	29	122	58	2,374,401	291	121	58	2,376,899
					292	1	1	2,074,395
	31	1365	129	1,615,539	311	225	59	1,707,057
					312	54	16	1,792,685
					314	187	39	1,755,860
					315	283	99	1,319,459
					316	503	86	1,822,314
					317	113	39	1,169,108
	32	3	3	1,611,092	321	2	2	1,219,993
					323	1	1	1,808,087
	41	4	2	1,516,410	411	4	2	1,516,410
	42	100	46	1,035,048	421	89	42	1,014,510
					422	11	5	1,182,673
	44	3	1	1,389,947	444	3	1	1,006,307
	51	21	5	1,560,984	511	15	4	1,822,583
					512	6	1	1,202,050
	61	4	3	1,211,788	613	3	2	1,409,136
					616	1	1	855,122
	71	17	6	1,901,308	711	3	1	880,244
					712	8	4	1,376,672
					714	6	1	2,035,183
	72	4	3	1,735,334	721	1	1	1,915,889
					722	3	2	1,563,049
	73	266	41	2,017,683	731	11	4	1,894,817
					732	162	29	2,053,624
					733	93	23	1,988,068
	74	37	5	1,663,404	743	7	2	1,615,871
					744	30	3	1,704,409

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	75	2291	87	1,305,195	752	2	1	931,000
					753	2280	86	1,314,018
					754	9	2	1,248,099
	81	279	29	1,800,769	813	107	1	1,061,957
					815	82	9	2,021,481
					816	90	21	2,163,569
	82	3712	104	1,542,018	821	4	2	2,156,308
					822	39	2	1,539,175
					823	34	2	1,474,666
					825	12	2	1,445,020
					826	3619	100	1,381,368
					829	4	2	1,660,810
					83	51	8	1,502,634
	83	51	8	1,502,634	831	39	3	1,278,073
					832	5	3	1,731,471
					833	7	2	2,871,044
	84	157	48	1,726,993	842	147	46	1,549,710
					843	10	6	2,411,713
	91	102	42	1,128,452	911	20	13	1,154,723
					912	79	39	1,133,138
					914	3	3	799,996
	92	2	1	932,215	920	2	1	932,215
	93	598	60	1,209,552	930	598	60	1,209,552
	94	12	6	1,422,996	942	12	6	1,422,996
24	2	391	80	3,492,787	21	22	16	4,497,838
					23	150	67	3,307,481
					24	219	55	3,469,080
	3	37	32	2,899,015	30	37	32	2,899,015
	11	125	6	2,117,639	112	124	6	2,063,020
					113	1	1	2,833,050
	12	10	4	2,713,358	120	10	4	2,713,358
	13	318	41	2,785,669	131	11	3	4,090,549
					132	58	11	2,757,665
					133	233	35	2,621,669
					135	16	11	3,570,849
	14	67	21	1,703,468	141	2	1	2,764,791
					142	48	11	2,495,247
					143	12	9	1,843,073
145					5	5	1,210,780	

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	16	89	26	3,436,545	161	46	11	3,192,926
					162	10	7	3,299,688
					163	13	11	3,054,532
					164	1	1	4,320,965
					165	19	6	3,894,196
	18	2	1	1,192,442	183	2	1	1,198,112
	21	76	10	2,100,756	211	13	1	2,403,040
					212	63	10	2,066,639
	22	86	19	2,276,737	220	86	19	2,276,737
	23	1429	69	2,435,421	231	20	5	2,923,025
					232	179	26	3,046,108
					233	581	49	2,406,633
					234	29	11	2,128,614
					235	1	1	2,282,388
					237	594	48	2,322,712
					238	25	8	2,451,896
					241	2	2	1,595,115
	26	201	33	2,542,983	261	4	1	3,294,488
					262	193	32	2,541,354
					263	4	2	2,219,162
	29	306	69	2,578,530	291	300	68	2,577,660
					292	6	3	2,643,943
	31	2392	101	1,769,309	311	439	63	1,911,192
					312	242	32	2,154,911
					314	615	41	1,867,789
					315	330	85	1,415,551
					316	549	71	1,963,484
					317	214	42	1,325,764
					318	3	2	3,013,467
	32	28	15	1,567,007	321	7	7	1,259,332
					322	7	7	1,571,909
					323	14	2	1,798,474
	42	26	16	1,039,525	421	26	16	1,021,629
	44	9	7	2,178,322	442	1	1	2,471,511
					444	8	6	2,144,493
	51	598	9	2,005,290	511	291	6	2,040,008
					512	36	3	1,566,559
					513	271	2	2,372,150
	61	2	2	1,229,233	613	2	2	1,514,899

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	71	19	6	2,158,656	712	3	2	1,750,148
					713	11	2	2,511,716
					714	5	2	2,372,287
	72	12	6	1,839,388	721	10	5	1,959,106
					722	2	1	1,602,960
	73	625	41	2,516,679	731	11	4	2,066,021
					732	328	27	2,602,118
					733	286	32	2,470,622
	74	25	2	1,686,764	741	25	2	1,868,664
	75	574	6	1,385,858	751	84	2	1,302,704
					753	490	4	1,440,761
	81	1613	45	2,451,570	812	24	4	1,409,424
					815	1342	34	2,655,789
					816	207	23	2,519,902
					817	40	2	1,910,511
	82	2068	53	1,665,457	821	7	2	2,372,567
					822	1274	41	1,722,204
					823	19	4	1,542,674
					824	1	1	2,995,685
					825	2	1	1,438,155
					826	642	5	1,557,780
					829	123	7	1,814,584
	83	25	5	1,836,651	831	12	3	1,720,532
					832	5	2	1,838,265
					833	8	1	2,746,806
	84	117	41	1,755,832	842	98	38	1,612,750
					843	19	6	2,413,399
91	70	26	1,170,317	911	8	6	1,130,806	
				912	35	20	1,151,931	
				913	26	2	1,591,573	
				914	1	1	1,050,814	
93	633	36	1,212,926	930	633	36	1,212,926	
94	5	2	1,501,437	942	5	2	1,501,437	
32	2	234	61	3,539,141	21	22	18	4,725,470
					23	103	46	3,318,679
					24	109	47	3,494,696
	3	49	36	3,108,145	30	49	36	3,108,145
	11	5	1	2,186,539	113	5	1	3,361,793
	12	166	14	2,837,011	120	166	14	2,837,011

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	13	476	33	2,735,269	131	3	3	3,788,737
					132	402	31	2,607,585
					133	28	3	2,702,237
					135	43	7	3,511,918
	14	6	6	1,670,400	143	2	2	1,872,507
					145	4	4	1,209,950
	16	66	21	3,413,465	161	21	11	3,311,492
					162	10	8	3,470,053
					163	9	7	2,972,581
					164	7	1	3,810,470
					165	19	8	3,657,727
	18	5	2	1,210,100	183	5	2	1,215,208
	21	4	2	2,113,087	211	2	1	2,330,975
					213	2	1	3,057,100
	22	224	26	2,193,807	220	224	26	2,193,807
	23	1682	63	2,263,587	231	2	2	2,866,469
					232	948	40	2,394,590
					233	65	8	2,357,438
					234	77	16	1,861,449
					235	11	2	2,525,692
					237	472	49	2,116,850
					238	107	9	2,397,388
					24	2	2	1,656,204
	26	77	28	2,571,806	261	6	3	3,330,184
					262	70	27	2,556,322
					263	1	1	2,413,353
	29	224	58	2,606,122	291	224	58	2,605,798
	31	1980	88	1,802,678	311	248	49	1,905,842
					312	248	19	2,240,640
					314	394	54	1,865,799
					315	280	71	1,504,934
					316	655	58	1,944,978
317					153	38	1,362,750	
318					2	1	1,919,732	
32	9	8	1,538,316	321	2	2	1,298,154	
				322	6	5	1,393,205	
				323	1	1	1,828,581	
41	3	3	1,726,630	411	3	3	1,726,630	
42	26	12	1,038,572	421	26	12	1,022,607	

산업 중분류	직업 중분류	근로자수	사업체수	가중평균임금	직업 소분류	근로자수	사업체수	가중평균임금
	44	38	4	2,702,411	442	33	2	2,910,108
					444	5	2	1,928,507
	51	2	1	2,008,709	511	2	1	2,046,433
					61	2	2	1,569,379
	71	10	3	2,038,692	713	9	2	2,442,447
					714	1	1	1,669,181
	72	106	10	1,602,986	721	94	8	1,592,752
					722	12	2	1,687,258
	73	1062	39	2,295,498	731	74	3	1,902,630
					732	113	14	2,462,944
					733	875	34	2,258,648
	74	33	6	1,572,967	741	24	3	1,316,364
					744	9	3	1,731,449
	75	1	1	1,386,025	752	1	1	1,068,299
	81	325	17	2,351,051	812	3	2	1,394,699
					815	5	2	2,644,126
					816	113	9	2,450,749
					817	204	8	1,636,166
	82	711	25	1,632,955	821	129	12	1,687,686
					822	183	12	1,658,429
					823	38	4	1,381,586
					825	13	3	1,465,880
					829	348	9	1,508,584
	83	3705	51	1,554,299	831	257	10	1,414,267
					832	3445	46	1,566,062
					833	3	1	1,284,516
	84	42	23	1,762,990	842	37	22	1,638,286
					843	5	4	2,393,718
	91	48	19	1,190,143	911	17	10	1,168,288
912					26	13	1,160,692	
913					1	1	1,604,765	
914					4	1	1,322,032	
93	770	30	1,184,275	930	770	30	1,184,275	
94	1	1	1,546,480	942	1	1	1,546,480	

3.2 현행 임금구조통계조사에 대한 평가

(1) 산업소분류별 통계 작성에 대한 검토

현행 표본설계는 각 산업중분류 단위에서 목표오차를 관리하고 있다. 또한 산업중분류 내의 산업소분류 수에 따라서 목표오차에 차등을 두고 있고, 사업체 수가 적은 산업중분류에 대해서는 전수조사를 하기 때문에 산업소분류 단위의 통계는 상당히 안정적인 것으로 예상된다. 추가적인 분석은 2002년 임금구조기본통계조사 결과가 정리된 이후에 가능하다.

(2) 직종세분류별 노동통계 작성에 대한 검토

2002년부터 조사되기 시작한 새로운 표본설계는 직종분류에 대해서 과거 조사에 비해서 안정적인 추정이 가능하다. 그러나 표본설계 단계에서 직종세분류별 추정을 반영하는 것은 불가능하기 때문에 직종세분류별 추정의 정확도는 그 편차가 크게 된다. 한편, 2002년 조사결과는 근로자의 직종소분류 구분까지만 조사하고 있기 때문에 직종세분류에 대한 노동통계 현황에 대해서는 알 수 없는 형편이다. 2003년 임금구조기본통계조사에서는 부분적으로 근로자의 직종을 세분류까지 조사하였다. 2003년 조사결과가 정리되면 직종세분류에 대한 평균임금을 소지역통계 기법을 적용해서 작성할 수 있을 것이다.

(3) 소지역추정법을 활용하는 방안

일반적으로 소지역 통계를 작성할 때 소지역은 계획된 관심영역과 계획되지 않은 관심영역으로 구분된다.

① 계획된 관심영역(planned domain)

조사설계 단계에서 해당 영역에 대한 추정을 염두에 두고 표본설계에 반영된 관심영역을 말한다. 일반적으로 층(strata)에 해당한다. 만약 우리 나라 경제활동인구조사에 시도별 추정을 위해서 표본설계 단계에서 각 시.도를 층으로 구분하였다면 각 시.도는 계획된 관심영역에 해당한다. 임금구조기본통계조사에서 산업중분류는 계획된 관심영역에 해당한다.

② 계획되지 않은 관심영역(unplanned domain)

표본설계 단계에서 해당 관심영역에 대한 추정을 고려되지 않은 경우이다. 예를 들어 경제활동인구조사에서 각 시.도별 통계 작성을 목적으로 표본설계되었는데, 각 시.군.구에 대한 소지역 통계 작성이 요구된다면, 이 경우에 각 시.군.구는 계획되지 않은 관심영역이 된다. 임금구조기본통계조사에서 직종분류는 계획되지 않은 관심영역에 해당한다.

산업소분류나 직종세분류 구분의 노동통계를 작성할 때 표본조사를 통해서 얻어진 정보에만 의존하게 되면 추정오차가 커지게 되어 신뢰성 있는 추정값을 얻을 수 없다. 소지역 통계 생산을 위해 무작정 표본 사업체를 증가시키는 것은 조사비용이 늘고, 표본 관리에 어려움이 많다. 소지역 통계 작성할 수 있는 범위는 다음과 같다.

- 산업소분류별 근로자 수, 평균 임금에 대한 통계
- 직종세분류별 근로자 수, 평균 임금에 대한 통계

다만, 회귀 산업, 직종에 대한 처리 : 모집단 사업체 및 근로자 수에 따라서 통계발표에 제한을 두거나 발표양식을 달리하는 방안 모색해야 할 것이다.

제 4 장 산업소분류 평균임금 산출을 위한 소지역 추정기법

효율적인 소지역 추정기법을 개발하기 위해서는 표본조사 자료에서 얻어지는 산업소분류별 평균임금에 대한 직접추정량(direct estimator)의 신뢰성을 보완할 수 있는 효과적인 간접추정량(indirect estimator)을 개발하는 것이 필요하다. 또한 모형기반 접근법인 혼합선형모형(mixed linear model)을 이용한 EBLUP(empirical best linear unbiased predictor) 추정기법에 대한 연구가 요구된다. 이를 위해서 본 연구에서는 다음과 같은 과정을 통해 효율적인 소지역 추정기법을 구현하고자 한다.

- 관심변수인 근로자의 평균임금을 효율적으로 설명할 수 있는 보조변수에 대한 분석
- 다른 통계 또는 전수조사에서 선택된 보조변수와 관련된 정보의 확보 가능성 검토
- 선택된 보조변수를 효과적으로 활용한 선형모형을 이용한 간접추정방법의 구현
- 직접추정량과 간접추정량을 활용한 효율적인 복합추정량(composite estimator)의 구현
- 적절한 혼합선형모형을 이용한 EBLUP 추정법의 구현

우선 근로자의 임금을 효과적으로 설명할 수 있는 보조변수로 어떤 것들이 있는지 파악하기 위해 임금구조통계조사 자료를 기초로 상관분석, 결정나무분석, 회귀분석을 순차적으로 실시하고자 한다.

▶ 기초분석대상 업종(산업)

소지역 모형개발을 위한 기초적인 분석에 있어서는 다수의 모든 산업분류를 동시에 고려하는 것은 현실적으로 어렵기 때문에 제조업을 주 분석대상으로 하고 있으며, 동시에 산업중분류에 대한 분석을 필요로 하는 경우 제조업내의 산업중분류 중 다음 4개 중분류를 주요 분석 대상으로 한다. 여기서 이들 중분류를 선택한 이유는 기초적인 소지역 추정기법을 구현하고 이에 대한 효율성 검증을 위해서는 어느 정도 충분한 크기의 표본을 확보할 수 있는 경우를 분석대상으로 하는 것이 필요하다는 점을 고려한 것이다.

- 중분류 15 : 음식료품 제조업
- 중분류 17 : 섬유제품 제조업
- 중분류 24 : 화합물 및 화학제품 제조업
- 중분류 32 : 영상, 음향 및 통신장비 제조업

한편 이들 제조업 중분류에 대한 기초분석을 통해 최종 소지역 모형이 선정된 후 실제 산업 소분류별 평균임금 산출에 있어서는 위에 제시된 제조업의 중분류에 추가로 도소매업의 일부 산업소분류도 분석대상에 포함한다.

4.1 상관분석

관심대상 변수인 근로자 임금총액을 설명할 수 있는 변수의 선정을 위해 우선 제조업 내 모든 중분류에 대해 상관분석을 실시하였고, 아울러 제조업내 중분류 중 15, 17, 24, 32에 속하는 모든 소분류에 대한 상관분석을 하였다. 실제 분석에 있어서는 관심변수로 정액급여, 임금총액, 그리고 이들 정액급여와 임금총액을 로그변환한 경우 등 모두 4개의 변수들을 대상으로 다른 변수들과의 상관분석을 해 본 결과 임금총액과 로그변환한 임금총액이 근속년수 등과 같은 다른 설명변수들과의 상관관계가 상대적으로 높은 것으로 나타난다. 로그변환한 임금총액의 경우 임금총액을 사용하는 경우와 큰 차이가 없고, 로그변환을 취하는 경우 소지역 모형개발에 있어서 변수변환에 따른 이론적으로 복잡한 문제가 발생하기 때문에 본 연구의 1차년도 소지역 모형개발에서는 노동자 임금을 나타내는 변수로 “임금총액”을 사용한다. 로그변환을 취한 임금총액을 활용하는 소지역 모형개발 및 효율성 검토에 대해서는 향후 좀더 심층적인 분석이 필요하다고 판단된다.

실제 분석결과 중 보고서에는 제조업내 중분류별 분석결과만을 <표 4.1>에 수록하였다. 상관분석결과를 보면 중분류에 따라 상당한 차이를 보이는 경우도 나타나고 있지만 대체적으로 근속년수(또는 경력년수), 나이, 사업체규모, 기업체규모 등이 임금총액과 상당히 높은 상관관계를 보여 주고 있음을 볼 수 있다.

참고로 상관분석은 연속형 변수들의 관계를 규명하기 위한 분석기법이다. 하지만 임금구조통계조사에서는 경력년수, 사업체규모, 기업체규모 등은 범주형 변수로 조사되고 있다. 따라서 이들 주요 범주형 변수에 대해 직접적으로 상관분석을 적용할 수 없기 때문에 여기서는 이들 3개 범주형 변수에 대해서는 각 범주의 중간값을 관측값으로 간주하고 분석한 것이다. 따라서 이런 불가피한 상황으로 인해 실제 이들과 임금총액과의 상관계수는 <표 4.1>에 나타난 상관계수에 비해 클 것으로 예상된다.

▶ 의사결정나무 분석의 활용

상관분석의 경우 범주형 변수를 연속형 변수로 변환하여 사용해야 한다는 한계를 갖고 있기 때문에 범주형 변수 및 연속형 변수를 모두 그대로 사용하여 변수들간의 연관관계를 규명할 수 있는 의사결정나무(decision tree)분석을 실시하였다. 임금총액을 목표(target)변수로 하고 임금구조통계조사의 나머지 변수들을 입력(input)변수로 설정하여 분석한 결과 임금결정에 영향을 주는 주요변수로는 근속년수(또는 경력년수), 직급, 성별, 직업소분류, 기업체규모 등인 것으로 나타났다. 의사결정나무 기법에 의한 분석결과는 <표 4.1>의 상관분석결과 및 다음에 제시될 회귀분석을 이용한 변수선택결과와 큰 차이를 보이지 않기 때문에 이에 대한 구체적인 분석결과는 보고서에 수록하지 않았다.

<표 4.1> 제조업 중분류별 상관분석 (총임금)

(1) 음식료품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.159	0.502	0.502	-0.054	0.169	0.189	0.457
나이	0.159	1.000	0.445	0.444	0.035	-0.072	-0.067	0.456
근속년수	0.502	0.445	1.000	0.999	-0.049	0.224	0.230	0.804
근속월수	0.502	0.444	0.999	1.000	-0.049	0.225	0.231	0.804
정상근로일수	-0.054	0.035	-0.049	-0.049	1.000	-0.055	-0.140	-0.050
사업체규모	0.169	-0.072	0.224	0.225	-0.055	1.000	0.529	0.131
기업체규모	0.189	-0.067	0.230	0.231	-0.140	0.529	1.000	0.139
경력년수	0.457	0.456	0.804	0.804	-0.050	0.131	0.139	1.000

(2) 담배제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.301	0.609	0.606	0.460	0.066	-0.343	0.504
나이	0.301	1.000	0.782	0.782	-0.041	-0.061	0.049	0.578
근속년수	0.609	0.782	1.000	1.000	0.283	-0.033	-0.251	0.755
근속월수	0.606	0.782	1.000	1.000	0.282	-0.032	-0.250	0.755
정상근로일수	0.460	-0.041	0.283	0.282	1.000	-0.291	-0.646	0.273
사업체규모	0.066	-0.061	-0.033	-0.032	-0.291	1.000	0.302	0.057
기업체규모	-0.343	0.049	-0.251	-0.250	-0.646	0.302	1.000	-0.259
경력년수	0.504	0.578	0.755	0.755	0.273	0.057	-0.259	1.000

(3) 섬유제품 제조업 : 봉제의복 제외

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.209	0.575	0.575	-0.094	0.162	0.100	0.515
나이	0.209	1.000	0.439	0.437	0.037	-0.142	-0.175	0.504
근속년수	0.575	0.439	1.000	0.999	-0.080	0.299	0.180	0.737
근속월수	0.575	0.437	0.999	1.000	-0.080	0.302	0.181	0.737
정상근로일수	-0.094	0.037	-0.080	-0.080	1.000	-0.104	0.042	0.009
사업체규모	0.162	-0.142	0.299	0.302	-0.104	1.000	0.605	0.098
기업체규모	0.100	-0.175	0.180	0.181	0.042	0.605	1.000	0.068
경력년수	0.515	0.504	0.737	0.737	0.009	0.098	0.068	1.000

(4) 봉제의복 및 모피제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	-0.059	0.290	0.290	-0.253	0.232	0.242	0.287
나이	-0.059	1.000	0.336	0.336	0.277	-0.223	-0.140	0.426
근속년수	0.290	0.336	1.000	0.998	0.089	0.177	0.145	0.652
근속월수	0.290	0.336	0.998	1.000	0.091	0.176	0.144	0.652
정상근로일수	-0.253	0.277	0.089	0.091	1.000	-0.465	-0.406	0.080
사업체규모	0.232	-0.223	0.177	0.176	-0.465	1.000	0.669	0.037
기업체규모	0.242	-0.140	0.145	0.144	-0.406	0.669	1.000	0.029
경력년수	0.287	0.426	0.652	0.652	0.080	0.037	0.029	1.000

(5) 가죽, 가방 및 신발 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.120	0.373	0.373	-0.155	0.157	0.167	0.317
나이	0.120	1.000	0.391	0.389	-0.045	-0.141	-0.202	0.377
근속년수	0.373	0.391	1.000	0.999	-0.286	0.256	0.162	0.686
근속월수	0.373	0.389	0.999	1.000	-0.287	0.258	0.164	0.685
정상근로일수	-0.155	-0.045	-0.286	-0.287	1.000	-0.448	-0.431	-0.113
사업체규모	0.157	-0.141	0.256	0.258	-0.448	1.000	0.451	0.188
기업체규모	0.167	-0.202	0.162	0.164	-0.431	0.451	1.000	0.160
경력년수	0.317	0.377	0.686	0.685	-0.113	0.188	0.160	1.000

(6) 목재 및 나무제품 제조업 : 가구제외

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.125	0.429	0.431	-0.073	0.248	0.275	0.433
나이	0.125	1.000	0.430	0.430	0.050	0.017	-0.084	0.481
근속년수	0.429	0.430	1.000	0.999	-0.153	0.277	0.062	0.759
근속월수	0.431	0.430	0.999	1.000	-0.154	0.280	0.064	0.759
정상근로일수	-0.073	0.050	-0.153	-0.154	1.000	-0.232	-0.148	-0.082
사업체규모	0.248	0.017	0.277	0.280	-0.232	1.000	0.556	0.091
기업체규모	0.275	-0.084	0.062	0.064	-0.148	0.556	1.000	-0.009
경력년수	0.433	0.481	0.759	0.759	-0.082	0.091	-0.009	1.000

(7) 펄프, 종이 및 종이제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.281	0.603	0.603	-0.155	0.463	0.157	0.553
나이	0.281	1.000	0.509	0.508	0.103	-0.095	0.049	0.457
근속년수	0.603	0.509	1.000	0.999	-0.126	0.403	0.238	0.789
근속월수	0.603	0.508	0.999	1.000	-0.127	0.404	0.239	0.790
정상근로일수	-0.155	0.103	-0.126	-0.127	1.000	-0.216	-0.365	-0.147
사업체규모	0.463	-0.095	0.403	0.404	-0.216	1.000	0.334	0.327
기업체규모	0.157	0.049	0.238	0.239	-0.365	0.334	1.000	0.240
경력년수	0.553	0.457	0.789	0.790	-0.147	0.327	0.240	1.000

(8) 출판, 인쇄 및 기록매체 복제업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.375	0.549	0.549	-0.111	0.418	0.205	0.465
나이	0.375	1.000	0.534	0.535	0.072	0.087	-0.010	0.565
근속년수	0.549	0.534	1.000	0.999	-0.047	0.383	0.235	0.763
근속월수	0.549	0.535	0.999	1.000	-0.044	0.382	0.233	0.766
정상근로일수	-0.111	0.072	-0.047	-0.044	1.000	-0.321	-0.573	0.034
사업체규모	0.418	0.087	0.383	0.382	-0.321	1.000	0.676	0.249
기업체규모	0.205	-0.010	0.235	0.233	-0.573	0.676	1.000	0.116
경력년수	0.465	0.565	0.763	0.766	0.034	0.249	0.116	1.000

(9) 코크스, 석유정제품 및 핵연료 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.489	0.637	0.637	-0.211	0.406	0.403	0.511
나이	0.489	1.000	0.651	0.651	-0.004	0.010	0.001	0.533
근속년수	0.637	0.651	1.000	0.999	-0.078	0.370	0.218	0.802
근속월수	0.637	0.651	0.999	1.000	-0.079	0.372	0.219	0.803
정상근로일수	-0.211	-0.004	-0.078	-0.079	1.000	-0.122	-0.530	0.002
사업체규모	0.406	0.010	0.370	0.372	-0.122	1.000	0.510	0.369
기업체규모	0.403	0.001	0.218	0.219	-0.530	0.510	1.000	0.143
경력년수	0.511	0.533	0.802	0.803	0.002	0.369	0.143	1.000

(10) 화합물 및 화학제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.583	0.666	0.667	-0.028	0.336	-0.010	0.641
나이	0.583	1.000	0.708	0.708	0.068	0.010	-0.160	0.641
근속년수	0.666	0.708	1.000	0.999	0.062	0.271	-0.010	0.850
근속월수	0.667	0.708	0.999	1.000	0.063	0.272	-0.009	0.850
정상근로일수	-0.028	0.068	0.062	0.063	1.000	0.092	0.027	0.040
사업체규모	0.336	0.010	0.271	0.272	0.092	1.000	0.403	0.258
기업체규모	-0.010	-0.160	-0.010	-0.009	0.027	0.403	1.000	-0.012
경력년수	0.641	0.641	0.850	0.850	0.040	0.258	-0.012	1.000

(11) 고무 및 플라스틱제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.192	0.629	0.630	-0.212	0.244	0.260	0.592
나이	0.192	1.000	0.502	0.502	0.056	0.067	-0.008	0.448
근속년수	0.629	0.502	1.000	0.999	-0.175	0.421	0.277	0.801
근속월수	0.630	0.502	0.999	1.000	-0.175	0.422	0.278	0.801
정상근로일수	-0.212	0.056	-0.175	-0.175	1.000	-0.269	-0.281	-0.148
사업체규모	0.244	0.067	0.421	0.422	-0.269	1.000	0.668	0.395
기업체규모	0.260	-0.008	0.277	0.278	-0.281	0.668	1.000	0.392
경력년수	0.592	0.448	0.801	0.801	-0.148	0.395	0.392	1.000

(12) 비금속광물제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.256	0.551	0.549	0.027	0.425	0.377	0.511
나이	0.256	1.000	0.559	0.559	0.006	-0.039	-0.014	0.461
근속년수	0.551	0.559	1.000	0.999	0.009	0.350	0.311	0.789
근속월수	0.549	0.559	0.999	1.000	0.008	0.350	0.310	0.790
정상근로일수	0.027	0.006	0.009	0.008	1.000	0.029	0.248	0.027
사업체규모	0.425	-0.039	0.350	0.350	0.029	1.000	0.525	0.387
기업체규모	0.377	-0.014	0.311	0.310	0.248	0.525	1.000	0.297
경력년수	0.511	0.461	0.789	0.790	0.027	0.387	0.297	1.000

(13) 제 1차 금속산업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.393	0.538	0.539	0.006	0.260	0.095	0.520
나이	0.393	1.000	0.581	0.581	0.044	0.089	-0.043	0.504
근속년수	0.538	0.581	1.000	0.999	-0.002	0.446	0.158	0.823
근속월수	0.539	0.581	0.999	1.000	-0.002	0.447	0.158	0.824
정상근로일수	0.006	0.044	-0.002	-0.002	1.000	0.027	-0.087	-0.001
사업체규모	0.260	0.089	0.446	0.447	0.027	1.000	0.330	0.371
기업체규모	0.095	-0.043	0.158	0.158	-0.087	0.330	1.000	0.090
경력년수	0.520	0.504	0.823	0.824	-0.001	0.371	0.090	1.000

(14) 조립금속제품 제조업; 기계 및 가구 제외

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.217	0.495	0.496	-0.254	0.434	0.490	0.505
나이	0.217	1.000	0.464	0.465	0.007	-0.032	0.126	0.465
근속년수	0.495	0.464	1.000	0.999	-0.235	0.279	0.359	0.796
근속월수	0.496	0.465	0.999	1.000	-0.235	0.280	0.360	0.796
정상근로일수	-0.254	0.007	-0.235	-0.235	1.000	-0.449	-0.348	-0.166
사업체규모	0.434	-0.032	0.279	0.280	-0.449	1.000	0.574	0.275
기업체규모	0.490	0.126	0.359	0.360	-0.348	0.574	1.000	0.240
경력년수	0.505	0.465	0.796	0.796	-0.166	0.275	0.240	1.000

(15) 기타 기계 및 장비 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.430	0.580	0.581	-0.003	0.425	0.316	0.573
나이	0.430	1.000	0.571	0.571	0.085	0.067	0.060	0.568
근속년수	0.580	0.571	1.000	0.999	-0.015	0.446	0.365	0.823
근속월수	0.581	0.571	0.999	1.000	-0.015	0.447	0.365	0.823
정상근로일수	-0.003	0.085	-0.015	-0.015	1.000	-0.102	0.039	-0.016
사업체규모	0.425	0.067	0.446	0.447	-0.102	1.000	0.630	0.385
기업체규모	0.316	0.060	0.365	0.365	0.039	0.630	1.000	0.332
경력년수	0.573	0.568	0.823	0.823	-0.016	0.385	0.332	1.000

(16) 컴퓨터 및 사무용 기기 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.395	0.544	0.544	-0.136	0.074	0.103	0.618
나이	0.395	1.000	0.453	0.453	-0.047	-0.119	-0.184	0.509
근속년수	0.544	0.453	1.000	0.998	-0.201	0.098	0.006	0.790
근속월수	0.544	0.453	0.998	1.000	-0.197	0.096	0.003	0.790
정상근로일수	-0.136	-0.047	-0.201	-0.197	1.000	0.059	0.125	-0.153
사업체규모	0.074	-0.119	0.098	0.096	0.059	1.000	0.314	-0.055
기업체규모	0.103	-0.184	0.006	0.003	0.125	0.314	1.000	-0.050
경력년수	0.618	0.509	0.790	0.790	-0.153	-0.055	-0.050	1.000

(17) 기타 전기기계 및 전기변환장치 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.390	0.555	0.556	-0.004	0.362	0.339	0.544
나이	0.390	1.000	0.568	0.568	0.027	0.072	0.149	0.555
근속년수	0.555	0.568	1.000	0.999	-0.051	0.508	0.482	0.805
근속월수	0.556	0.568	0.999	1.000	-0.053	0.509	0.482	0.806
정상근로일수	-0.004	0.027	-0.051	-0.053	1.000	-0.228	0.100	-0.073
사업체규모	0.362	0.072	0.508	0.509	-0.228	1.000	0.688	0.394
기업체규모	0.339	0.149	0.482	0.482	0.100	0.688	1.000	0.348
경력년수	0.544	0.555	0.805	0.806	-0.073	0.394	0.348	1.000

(18) 전자부품, 영상, 음향 및 통신장비 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.412	0.573	0.574	0.021	0.298	0.177	0.590
나이	0.412	1.000	0.504	0.504	0.061	-0.155	0.023	0.529
근속년수	0.573	0.504	1.000	0.999	0.068	0.294	0.281	0.849
근속월수	0.574	0.504	0.999	1.000	0.068	0.296	0.280	0.849
정상근로일수	0.021	0.061	0.068	0.068	1.000	-0.031	0.255	0.025
사업체규모	0.298	-0.155	0.294	0.296	-0.031	1.000	0.387	0.222
기업체규모	0.177	0.023	0.281	0.280	0.255	0.387	1.000	0.218
경력년수	0.590	0.529	0.849	0.849	0.025	0.222	0.218	1.000

(19) 의료, 정밀, 광학기기 및 시계 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.178	0.449	0.451	-0.205	0.322	0.241	0.487
나이	0.178	1.000	0.398	0.397	-0.065	-0.094	-0.043	0.428
근속년수	0.449	0.398	1.000	0.999	-0.312	0.324	0.336	0.802
근속월수	0.451	0.397	0.999	1.000	-0.312	0.324	0.337	0.802
정상근로일수	-0.205	-0.065	-0.312	-0.312	1.000	-0.333	-0.073	-0.206
사업체규모	0.322	-0.094	0.324	0.324	-0.333	1.000	0.794	0.250
기업체규모	0.241	-0.043	0.336	0.337	-0.073	0.794	1.000	0.282
경력년수	0.487	0.428	0.802	0.802	-0.206	0.250	0.282	1.000

(20) 자동차 및 트레일러 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.324	0.594	0.595	0.039	0.488	0.098	0.564
나이	0.324	1.000	0.564	0.564	-0.050	0.022	0.039	0.550
근속년수	0.594	0.564	1.000	0.999	-0.134	0.456	0.177	0.860
근속월수	0.595	0.564	0.999	1.000	-0.133	0.458	0.179	0.860
정상근로일수	0.039	-0.050	-0.134	-0.133	1.000	-0.112	-0.352	-0.147
사업체규모	0.488	0.022	0.456	0.458	-0.112	1.000	0.394	0.389
기업체규모	0.098	0.039	0.177	0.179	-0.352	0.394	1.000	0.193
경력년수	0.564	0.550	0.860	0.860	-0.147	0.389	0.193	1.000

(21) 기타 운송장비 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.477	0.485	0.485	-0.189	0.323	0.167	0.450
나이	0.477	1.000	0.667	0.667	-0.095	0.109	0.014	0.606
근속년수	0.485	0.667	1.000	0.999	-0.232	0.438	0.173	0.829
근속월수	0.485	0.667	0.999	1.000	-0.232	0.438	0.172	0.830
정상근로일수	-0.189	-0.095	-0.232	-0.232	1.000	-0.377	-0.076	-0.178
사업체규모	0.323	0.109	0.438	0.438	-0.377	1.000	0.330	0.367
기업체규모	0.167	0.014	0.173	0.172	-0.076	0.330	1.000	0.175
경력년수	0.450	0.606	0.829	0.830	-0.178	0.367	0.175	1.000

(22) 가구 및 기타 제품 제조업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	0.255	0.519	0.521	-0.172	0.347	0.244	0.535
나이	0.255	1.000	0.528	0.530	-0.041	0.096	0.229	0.529
근속년수	0.519	0.528	1.000	0.999	-0.239	0.478	0.515	0.880
근속월수	0.521	0.530	0.999	1.000	-0.241	0.477	0.515	0.881
정상근로일수	-0.172	-0.041	-0.239	-0.241	1.000	-0.383	-0.366	-0.232
사업체규모	0.347	0.096	0.478	0.477	-0.383	1.000	0.697	0.413
기업체규모	0.244	0.229	0.515	0.515	-0.366	0.697	1.000	0.422
경력년수	0.535	0.529	0.880	0.881	-0.232	0.413	0.422	1.000

(23) 재생용 가공원료 생산업

	총임금	나이	근속년수	근속월수	정상근로일수	사업체규모	기업체규모	경력년수
총임금	1.000	-0.043	0.344	0.345	-0.135	0.233	0.157	0.328
나이	-0.043	1.000	0.182	0.184	0.129	-0.165	0.007	0.157
근속년수	0.344	0.182	1.000	0.998	-0.195	0.231	0.334	0.653
근속월수	0.345	0.184	0.998	1.000	-0.196	0.230	0.340	0.654
정상근로일수	-0.135	0.129	-0.195	-0.196	1.000	-0.541	-0.141	-0.064
사업체규모	0.233	-0.165	0.231	0.230	-0.541	1.000	-0.065	0.067
기업체규모	0.157	0.007	0.334	0.340	-0.141	-0.065	1.000	0.249
경력년수	0.328	0.157	0.653	0.654	-0.064	0.067	0.249	1.000

4.2 선형모형 구현을 위한 설명변수 선택

임금구조통계조사에서 조사되는 임금을 설명할 수 있는 유의적인 변수를 선택하기 위해 회귀분석의 변수선택과정을 수행하였다. 변수선택과정은 제조업전체 및 중분류 15, 17, 24, 32에 대해 각각 수행되었으며 유의적으로 선택된 변수를 정리하면 <표 4.2>와 같다. 여기서 “0”으로 표시된 변수들이 선택된 변수들을 나타내고 있다. 예를 들어 ”기술기능 정도“와 같은 경우 중분류 17과 중분류 32에서는 유의적인 변수로 선택되지 않고 있다. 이 결과를 근거로 다수의 중분류에서 유의적으로 선택된 변수들만을 모형에 포함한 회귀분석을 통해 소지역 추정모형을 개발하고자 한다.

<표 4.2> 회귀분석 변수선택 결과

	제조업전체	15	17	24	32
teny(근속년수)	0	0	0	0	0
sex(성별)	0	0	0	0	0
ability(학력)	0	0	0	0	0
job_class(직급)	0	0	0	0	0
cat_b(사업체규모 범주)	0	0	0	0	0
c_range(경력년수 범주)	0	0	0	0	0
work(근무형태)	0	0	0	0	0
marriage(혼인상태)	0	0	0	0	0
job(직업소분류)	0	0	0	X	0
labor_yn(노조유무)	0	0	X	0	0
age(나이)	0	0	0	0	0
skill(기술기능정도)	0	0	X	0	X
industry(산업분류)	0	0	0	0	0
areal(지방노동관서번호)	0	0	0	0	0
cat_c(기업체규모범주)	0	0	0	X	0
employ(고용형태)	0	X	X	X	0

▶ 분석대상 설명변수

상관분석결과 및 <표 4.2>의 변수 선택결과에 따라 향후 분석에서는 다른 설명이 없으면 다음 변수들을 포함한 선형모형을 검토대상으로 한다.

종속변수 : 임금총액

설명변수 :

연속형: 근속년수, 나이

범주형:

직종분류 ...

성별 ① 남자 ② 여자

노조유무 ① 조직 ② 비조직

혼인상태 ① 미혼 ② 기혼

학력 ① 국졸이하 ② 중졸 ③ 고졸 ④ 전문대졸 ⑤ 대졸이상

직급 ① 임원 ② 부장 ③ 과장 ④ 계장 ⑤ 십장 ⑥ 조장 ⑦ 반장보 ⑧ 비직급

근무형태 ① 정상근로 ② 시간근로 ⑩ 교대없음 ⑫ 2교대 ⑬ 3교대 ⑭ 격일제 ⑯ 시간제

사업체 규모 ① 5-9 ② 10-29 ③ 30-99 ④ 100-299 ⑤ 300-499 ⑥ 500+

기업체 규모 ① 5-9 ② 10-29 ③ 30-99 ④ 100-299 ⑤ 300-499

⑥ 500-999 ⑦ 1000-4999 ⑧ 5000-9999 ⑨ 10000+

경력년수 ① 1년미만 ② 1-2년미만 ③ 2-3년미만 ④ 3-4년미만

⑤ 4-5년미만 ⑥ 5-10년미만 ⑦ 10년이상

4.3 다른 통계조사의 보조정보 활용 방안 검토

근로자 임금과 관련된 보조정보를 제공할 수 있는 통계조사로는 사업체노동실태조사 및 노동력수요동향조사 등을 검토해 볼 필요가 있는 데 이 중 위에 제시된 평균임금을 설명할 수 있는 변수들에 대한 산업소분류별 정보를 얻을 수 있는 통계조사로는 노동력수요동향조사를 고려해 볼 수 있다. 사업체노동실태조사는 사업체를 기본단위로 한 조사이기 때문에 개별 근로자에 대한 정보를 얻을 수 없어서 활용에 한계가 있다.

한편 노동력수요동향조사의 경우 앞에서 변수선택과정을 통해 선택된 설명변수 중 종사자수, 성별구성비율, 직종에 대한 정보만이 활용이 가능하다. 따라서 노동력수요동향조사결과를 보조정보로 활용하는 것을 전제로 회귀분석을 실시하여 모형의 적합성을 나타내는 결정계수(R^2)를 구하면 <표 4.3>과 같다. <표 4.3>의 결정계수들은 각각 제조업전체 및 제조업 중 산업중분류 15, 17, 24, 32의 경우에 회귀모형을 적합한 결과를 나타내며 각 산업분류에 대한 회귀모형에서 설명변수 중 직종분류를 대분류, 중분류 또는 소분류로 구분하여 사용하는 경우 얻어지는 15개 모형에 대한 분석 결과 얻어진 결정계수를 정리한 것이다.

<표 4.3> 노동력수요동향조사 활용에 따른 모형의 결정계수

	R^2		
	직업소분류	직업중분류	직업대분류
제조업전체	0.308838	0.281583	0.262714
15	0.375203	0.342388	0.319114
17	0.351586	0.302076	0.285377
24	0.371291	0.304156	0.280817
32	0.394724	0.368088	0.346695

※종속변수 : 임금총액

※설명변수 : 직종분류, 성별 ① 남자 ② 여자

<표 4.3>을 보면 검토된 모든 모형들의 결정계수는 상당히 낮게 나타나고 있어 노동력수요동향조사의 결과에서 얻어지는 보조정보만을 활용하는 모형은 실제 소지역 추정을 위한 활용에 있어 효율성에 문제가 있다고 판단된다. 따라서 산업소분류별 근로자 평균임금 추정을 위한 소지역 추정모형을 개발하는 데 있어서 임금구조통계조사 이외의 다른 조사에서 얻어지는 보조정보를 활용하는 것은 현재 상황에서는 한계가 있다고 판단된다. 따라서 현재 활용 가능한 소지역 추정모형은 임금구조통계조사의 자료에서 산업소분류를 산업중분류 또는 평균임금 등을 기준으로 군집화하여 모형을 개발하는 것이 최선의 방안이라고 판단된다.

4.4 소지역 추정을 위한 선형모형의 적합

산업소분류별 평균임금 추정에 활용될 수 있는 신뢰할 수 있는 보조정보를 다른 외부 통계를 통해 얻을 수 없다고 가정하는 경우 활용할 수 있는 최선의 간접추정방법은 임금 구조통계자료를 통해 일단 산업중분류별로 가장 효율적인 선형모형을 설정하고, 이 모형을 활용하여 얻어지는 산업소분류별 평균임금에 대한 예측값을 간접추정값으로 활용하는 것이 현실적으로 가장 효과적인 방안이라고 판단된다. 또 다른 방안은 산업소분류를 산업중분류 대신 평균임금을 기준으로 군집화하여 군집별로 모형을 적합시킨 후 이 모형을 활용하여 간접추정값을 산출하는 방안을 고려해 볼 수 있다.

이에 따라 우선 군집화 방안을 검토하기 위해 제조업내의 4개 중분류(15, 17, 24, 32)에 속하는 17개의 산업소분류를 산업중분류별로 군집화 하여 평균임금을 효과적으로 설명할 수 있는 모형을 검토하고, 아울러 분석대상 17개 산업소분류를 산업중분류 구분을 무시하고 각 소분류별 평균임금을 기준으로 유사한 소분류들을 하나의 군집으로 묶어 군집별로 모형을 적합하는 방안을 검토하여 그 결과를 비교하였다.

(1) 산업중분류 기준 선형모형의 검토

앞의 결과들을 살펴보면 평균임금에 대한 선형모형에서 직종분류는 상당히 의미 있는 설명변수인 것으로 판단된다. 모형에 직종분류를 반영하는 방안으로는 직종분류를 대분류, 중분류 또는 소분류 수준으로 사용하는 3가지 방법이 있을 수 있다. 여기서 직종분류를 대분류 또는 중분류 대신 소분류로 세분화하여 사용하는 경우 모형이 적합성이 상당 부분 향상될 수 있다면 소분류 수준에서 직종분류를 설명변수로 모형에 활용하는 것이 효과적이지만 이 경우 각 산업소분류에 속하는 자료에서 해당 직종 소분류에 속하는 관측수가 너무 적게 되면 산업소분류에 대한 평균임금 예측값의 신뢰성에 문제가 발생할 소지가 있기 때문에 어느 수준으로 직종분류를 활용하는 것이 적합한 것인지 검토해 볼 필요가 있다.

아울러 앞서도 언급된 것과 같이 임금구조통계조사에서는 중요한 설명변수인 경력년수, 사업체규모 및 기업체규모가 범주형으로 조사되고 있기 때문에 모형에서 이들 변수들을 범주형으로 반영할 것인지 또는 각 범주의 중간값으로 대체하여 반영할 것인지 검토해 볼 필요가 있다. 이에 따라 이들 가능한 모든 모형을 적합하여 모형의 적합성을 나타내는 결정계수를 정리해 보면 다음과 같다.

<표4.4> 중분류별 직종분류 및 범주형 변수 처리 방법에 따른 모형 결정계수

	직업소분류		직업중분류		직업대분류	
	중앙값	범주	중앙값	범주	중앙값	범주
제조업전체	0.569204	0.579128	0.555842	0.566176	0.550461	0.560528
15	0.548886	0.568383	0.528176	0.551522	0.518621	0.544113
17	0.623246	0.691878	0.605772	0.679670	0.599974	0.672156
24	0.689682	0.728129	0.678793	0.718706	0.669857	0.711853
32	0.679670	0.707250	0.670063	0.695888	0.667283	0.691994

<표 4.4>의 결과를 보면 직종분류의 적용에 있어서 소분류와 중분류를 사용하는 경우를 비교해 보면 실제 범주의 수에 있어서는 매우 큰 차이가 있음에도 불구하고 결정계수에 있어서는 큰 차이를 보이지 않고 있다. 아울러 범주형 변수의 처리에 있어서는 각 범주의 중간값을 사용하는 경우보다는 이들을 그대로 범주형 변수로 처리하는 것이 효율적이라는 것을 알 수 있다. 결과적으로 직종은 중분류 수준까지만 반영하고 범주로 조사된 변수의 경우 이를 중간값으로 환산하지 않고 그대로 범주형 변수로 처리하는 모형이 소지역 추정량 산출을 위한 가장 효과적인 모형이라는 결론을 내릴 수 있다.

▶ 산업소분류별 모형의 적합성 검토

한편 제시된 모형의 산업소분류별 적합성을 검토하기 위해 연구대상 중분류내의 산업소분류별로 모형을 적합시켜 얻어진 결정계수를 정리하면 다음 <표 4.5>와 같다.

<표 4.5> 산업소분류별 모형의 적합성 (R^2 및 root MSE)

		R^2		root MSE	
		직업소분류	직업중분류	직업소분류	직업중분류
15	151	0.709343	0.682574	421280.1	438173.3
	152	0.801358	0.789176	453792.4	462690.4
	153	0.665788	0.640772	644352.7	657442.8
	154	0.582004	0.559832	779068.1	796750.9
	155	0.672032	0.639604	685576.2	713540.4
17	171	0.760002	0.749236	464792.2	473765.1
	172	0.763311	0.743418	335432.8	346225.0
	173	0.835217	0.831918	324149.7	322311.2
	174	0.693171	0.686335	327693.9	329554.1
	179	0.756469	0.725237	370115.4	386367.5
24	241	0.762853	0.749238	685885.5	701604.2
	242	0.785205	0.774536	597732.8	609112.5
	243	0.774417	0.762516	497357.4	507687.2
	244	0.838274	0.833774	364870.6	367677.3
32	321	0.718727	0.692239	556606.3	580387.1
	322	0.734454	0.726247	569945.5	575093.4
	323	0.817740	0.810637	452708.6	459583.2

(2) 산업소분류별 평균임금 기준 군집 구성에 따른 모형 검토

산업소분류에 대한 로버스트한 선형모형을 구현하기 위해 앞에서 각 산업중분류내의 소분류를 하나의 유사 그룹으로 간주하고 각 중분류별로 모형을 적합하여 보았다. 이와 관련된 또 다른 접근방법은 각 산업소분류를 해당 중분류에 따라 분류하는 대신 소분류별 평균임금을 기준으로 평균임금이 유사한 산업소분류들을 하나의 그룹으로 구분하여 각 그룹별 모형을 적합시키는 방안을 생각해 볼 수 있다. 이런 방안의 효율성을 검토해 보기 위해 <표 4.5>에 제시된 17개 산업소분류를 산업중분류 대신 평균임금에 따라 다시 3개 그룹으로 재분류하고 각 그룹에 대해 모형을 적합한 결과를 요약하면 <표 4.6>과 같다. 한편 제조업 전체 산업소분류(69개)를 평균임금을 기준으로 5개씩 군집화하여 모형을 적합하여 그 적합도를 살펴보면 <표 4.7>과 같다.

<표 4.6> 평균임금 기준 군집화에 따른 모형 결정계수 (중분류 15, 17, 24, 32 대상)

	직업소분류	직업중분류
그룹1	0.661756	0.642611
그룹2	0.623360	0.610511
그룹3	0.700148	0.689474

* 그룹1 : 179, 172, 173, 151, 174 그룹2 : 171, 154, 321, 323, 322, 152, 243
 그룹3 : 153, 242, 155, 241, 244

<표 4.7> 평균임금 기준 군집화에 따른 모형 결정계수 (제조업전체 대상)

	직업소분류	직업중분류
그룹1	0.615781	0.605781
그룹2	0.660565	0.631614
그룹3	0.648506	0.631668
그룹4	0.632883	0.619269
그룹5	0.608031	0.588261
그룹6	0.594106	0.580525
그룹7	0.654669	0.635267
그룹8	0.684901	0.655755
그룹9	0.623056	0.607792
그룹10	0.728109	0.710014
그룹11	0.625127	0.609668
그룹12	0.563195	0.546585
그룹13	0.626573	0.560589
그룹14	0.700907	0.691880

그룹1: 193, 315, 179, 181, 172 그룹2: 173, 269, 151, 273, 331 그룹3: 372, 191, 174, 369, 334
 그룹3: 372, 191, 174, 369, 334 그룹4: 201, 361, 371, 192, 319 그룹5: 359, 171, 311, 222, 202
 그룹6: 289, 154, 231, 312, 300 그룹7: 333, 321, 323, 272, 262 그룹8: 292, 251, 293, 252, 291
 그룹9: 322, 271, 313, 332, 314 그룹10: 152, 295, 212, 243, 352 그룹11: 153, 341, 343, 353, 211
 그룹12: 242, 155, 263, 221, 351 그룹13: 261, 223, 233, 294, 241 그룹14: 232, 281, 244, 160

(3) 소지역 추정을 위한 최종 선형모형 선택

한편 4.2절에 제시된 설명변수들 중 일부는 모형에서 생략하여도 모형의 적합도에 크게 영향이 없고, 차후에 보조정보를 사용하거나 좀더 복잡한 형태의 모형을 사용할 수 있다는 점을 고려할 때, 가능하면 좀더 단순한 선형모형을 사용하는 것이 바람직하다. 이런 측면에서 제시된 설명변수 중 일부 변수를 생략하여 모형을 단순화시키는 경우 모형의 효율성을 검토해 본 결과는 <표 4.8>과 같다. <표 4.8>에서 완전모형은 4.2절에 제시된 모든 설명변수를 포함한 모형을 의미하며, 단순모형은 이중 나이, 혼인상태, 근무형태, 사업체규모는 생략한 모형을 의미한다.

<표 4.8> 모형 단순화에 따른 모형 결정계수

R^2		완전모형	단순모형
제조업	15	0.542675	0.544208
	17	0.630442	0.633521
	24	0.668332	0.672662
	32	0.659550	0.673215
도.소매업	50	0.363786	0.363804
	51	0.506690	0.508345
	52	0.660060	0.660231

<표 4.8>의 두 모형에 대한 비교 결과를 살펴보면 각 소분류에 있어서 큰 차이를 보이고 있지 않다는 것을 볼 수 있으며, 실제 생략된 나이, 사업체규모 등은 다른 설명변수로 충분히 설명될 수 있는 설명변수들이라는 것을 쉽게 파악할 수 있다. 이에 따라 노동부 관계자와의 협의를 통해 본 연구에서 소지역 추정 선형모형을 개발하기 위해 사용한 최종 설명변수는 다음과 같다.

연속형 : 근속년수

범주형 : 직종분류

성별 ① 남자 ② 여자

노조유무 ① 조직 ② 비조직

학력 ① 국졸이하 ② 중졸 ③ 고졸 ④ 전문대졸 ⑤ 대졸이상

직급 ① 임원 ② 부장 ③ 과장 ④ 계장 ⑤ 십장 ⑥ 조장 ⑦ 반장보 ⑧ 비직급

기업체 규모 ① 5-9 ② 10-29 ③ 30-99 ④ 100-299 ⑤ 300-499

⑥ 500-999 ⑦ 1000-4999 ⑧ 5000-9999 ⑨ 10000+

경력년수 ① 1년미만 ② 1-2년미만 ③ 2-3년미만 ④ 3-4년미만

⑤ 4-5년미만 ⑥ 5-10년미만 ⑦ 10년이상

(4) 간접추정량 산출

제시된 것과 같이 산업소분류의 군집화 방안으로 산업중분류를 이용하는 것과 평균임금을 기준으로 하는 방법을 비교해 보면, 모형의 적합도에 있어서 차이가 크지 않다는 것을 볼 수 있다. 이와 같이 두 가지 산업소분류 군집화 방법에 있어서 차이가 없다면 현재 일반적 산업분류 기준으로 널리 사용되고 있고, 동시에 군집화 기준에 대한 설명이 매우 명확한 산업중분류를 산업소분류 군집화 방안으로 활용하는 것일 효과적이라고 판단된다.

한편 일부 소규모 중분류(해당 사업체 수가 매우 적은 중분류)의 경우 로버스트한 모형을 적합시키는 데 문제가 발생할 수 있다. 이런 경우는 다른 유사한 소규모 중분류들을 하나의 그룹으로 묶어 처리하는 방법이 효과적일 것으로 예상된다. 따라서 이와 관련된 사항에 대해서는 차후 충분한 연구 검토작업이 필요할 것으로 생각된다.

적합된 산업중분류별 선형모형을 기초로 산업소분류별 평균임금에 대한 간접추정량을 산출한다. 이를 위해서는 앞에 제시된 바와 같이 우선 각 산업중분류별로 최적의 선형모형을 개발하고, 적합된 모형을 기초로 해당 소분류에 속하는 근로자들이 갖는 근속년수, 직종분류, 성별, 사업체규모 등의 특성을 설명변수에 반영하여 소분류별 평균임금에 대한 간접추정값을 산출한다. 이 과정에서 설명변수가 연속형인 경우 단순히 소분류별 해당 변수의 평균을 사용할 수 있지만, 제시된 모형에서와 같이 이들 설명변수들이 범주형인 경우 어떤 형태로 예측값을 산출할 것인가에 대해서는 향후 보다 면밀한 연구 검토가 필요하다. 본 연구에서 적용한 방법은 각 산업소분류에 속하는 중분류 수준에서 적합된 모형을 기초로 근로자 개인별 임금에 대한 예측값을 산출하고 이를 종합하여 소분류별 평균임금에 대한 간접추정량을 산출하는 방법을 사용하고 있다.

한편 모형에서 사용되는 설명변수에 대해 산업소분류별로 노동력수요동향조사 등 다른 통계조사를 통해 임금구조통계조사에서 얻을 수 있는 정보 보다 더 신뢰할 수 있는 정보를 얻을 수 있다면, 이들 정보를 부분적이라도 반영하는 방안에 대해서도 향후 검토가 필요하다.

4.5 산업소분류 평균임금 추정을 위한 복합추정

임금구조통계조사에서 얻어진 자료로부터 기존의 추정방법으로 직접 각 산업소분류별 평균임금을 산출하는 경우 일부 산업소분류의 경우 충분한 크기의 표본을 확보할 수 없기 때문에 추정량의 신뢰성을 확보할 수 없다. 또한 앞에서 제시한 것과 같이 동일 산업중분류에 속하는 다른 산업소분류의 유사정보를 이용하여 추정되는 간접추정량의 경우 잠재적인 편향(bias) 가능성이 항상 내재되어 있다. 따라서 산업소분류에 배정된 표본 수

가 적을 경우, 임금구조통계 자료만을 이용하여 추정되는 직접추정량의 불안정성과 간접추정량의 편향 가능성을 서로 보완할 수 있는 방법으로 두 추정량의 가중평균을 사용하는 방법을 생각할 수 있는데, 이렇게 두 추정량의 가중평균을 사용하는 것이 복합추정량 (composite estimator)이다. i 번째 산업소분류의 평균임금에 대한 복합추정식의 형태는 아래와 같다.

$$\widehat{Y}_{(COMP),i} = w_i \widehat{Y}_{1i} + (1-w_i) \widehat{Y}_{2i} \quad (4.1)$$

여기서 \widehat{Y}_{1i} 은 임금구조통계 조사자료에서 기존의 추정방법으로 얻은 i 번째 산업소분류의 근로자 평균임금에 대한 직접추정값이고, \widehat{Y}_{2i} 은 간접추정값으로 4.4절에 설명된 것과 같이 i 번째 산업소분류가 소속된 산업중분류에 포함된 모든 근로자의 정보를 활용하여 적합된 모형을 기반으로 산출된 i 번째 산업소분류 평균임금에 대한 간접추정값이다. 여기서 가중값 w_i 는 적당하게 선택된 0과 1사이의 값이다.

효율적으로 복합추정량을 활용하기 위해서는 적절한 가중값 w_i 를 결정하는 과정이 매우 중요하다. 이론적으로 가장 효율적인 방법은 복합추정량의 MSE를 최소로 하는 최적가중값을 찾는 방법이다. 이 경우 최적의 가중값은 다음과 같이 구해진다.

$$w_i(opt) = \frac{\widehat{MSE}(\widehat{Y}_{2i})}{\widehat{MSE}(\widehat{Y}_{2i}) + \widehat{Var}(\widehat{Y}_{1i})} \quad (4.2)$$

이런 최적 가중값을 사용하기 위해서는 각 산업소분류별로 직접추정량(\widehat{Y}_{1i})의 분산과 간접추정량(\widehat{Y}_{2i})의 평균제곱오차(MSE)를 추정하는 작업이 필요하게 되고, 이런 추정과정은 상당히 복잡한 과정을 수행해야하기 때문에 실제 적용상에 많은 어려움이 있다.

한편 최적가중값은 사용상 많은 어려움이 있기 때문에 최적가중값 외에 표본크기에 의해 결정하는 가중값이 사용되고 있다. 이런 가중값은 모집단 크기와 표본크기 또는 보조변수의 총계에 따라서만 가중값이 달라지게 된다. 현실적인 적용상의 편리성을 염두에 두고 본 연구에서는 다음과 같은 형태의 표본크기의존 추정량(sample-size dependent estimator)을 검토하고자 한다.

Drew, Singh과 Choudhry(1982)는 다음과 같은 가중값을 적용하는 표본크기의존 복합추정량을 제안하고 있다. i 번째 산업소분류의 모집단 크기, N_i 의 직접 불편추정량을 \widehat{N}_i 이라고 할 때,

$$w_i(D) = \begin{cases} 1 & , \widehat{N}_i \geq \delta N_i \text{ 일 때} \\ \frac{\widehat{N}_i}{\delta N_i} & , \text{ 그렇지 않을 때} \end{cases} \quad (4.3)$$

여기서 N_i 는 알고 있다고 가정한 값이고, δ 는 합성추정량의 기여도를 결정해 주는, 임의

로 선정되는 상수값이다. 예를 들어 Canadian Labour Force Survey의 경우 $\delta=2/3$ 를 사용하고 있다.

한편 Sarndal과 Hidiroglou(1989)는 (4.3)을 변형한 다음과 같은 가중값의 사용을 대안으로 제시하고 있다.

$$w_i(D) = \begin{cases} 1 & , \widehat{N}_i \geq N_i \text{ 일 때} \\ \left(\frac{\widehat{N}_i}{N_i} \right)^{h-1} & , \text{그렇지 않을 때} \end{cases} \quad (4.4)$$

여기서 h 는 임의로 선정되는 값이다. 이들은 일반적으로 $h=2$ 를 사용하는 것이 일반적인 경우 적절하다고 설명하고 있다. 여기서 $\delta=1$ 이고, $h=2$ 이면 (4.3)과 (4.4)는 동일해진다. 여기서 적절한 상수값 δ 또는 h 의 선택과 관련해서는 향후 추가적인 연구가 필요하다.

복합추정량을 산출하기 위해 (4.3)과 (4.4)를 적용하여 구한 복합추정결과는 <표 4.9>와 같다. <표 4.9>에 제시된 복합추정[1]은 (4.3)에서 $\delta=1$ 을 사용한 것이고, 복합추정량[2]은 (4.4)에서 $h=2.5$ 를 사용한 것이다. <표 4.9>에 제시된 두 가지 경우의 표본크기의존 추정량은 하나의 실례로 주어진 것이고 이들이 효율성에 대해서는 차후 추가적인 연구가 필요하다. 실제 효과적인 δ 또는 h 의 선택을 위해서는 이들 값의 선택에 따른 복합추정량의 효율성을 면밀히 검토해 보는 추가적인 연구가 필요하다.

<표 4.9>에서 직접추정값에 해당하는 평균임금은 조사 자료에서 직접구한 평균임금을 나타내며, 간접추정값에 해당하는 가중회귀결과는 앞에 제시된 선형모형을 통해 얻어진 각 소분류별 평균임금에 대한 추정값이다. 즉 직접추정값과 간접추정값은 각각 (4.1)에서 \widehat{Y}_{1i} 과 \widehat{Y}_{2i} 에 해당한다. 여기서 가중값이 “1”이 되는 경우는 각 소분류의 표본 사업체수가 비례배분에 위한 표본 사업체수에 비해 작지 않다고 판단된 것으로 이 경우 복합추정결과는 직접추정값에 해당하는 소분류 평균임금과 같아진 것이다.

한편 제시된 복합추정량의 정확도를 설명할 수 있는 평균제곱오차 $MSE(\widehat{Y}_{(COMP),i})$ 는 임금구조통계조사 자료로부터 추정되어야 하며 좀 더 안정적인 평균제곱오차에 대한 추정값들을 산출하기 위해 잭나이프 추정방법이 적용될 수 있다. 하지만 각 산업소분류 복합추정량에 대한 평균제곱오차를 잭나이프 기법으로 추정하는 작업은 그 과정이 매우 복잡하고 구체적인 프로그램을 별도로 작성해야 하는 현재적인 문제가 있기 때문에 이에 대한 대안으로 복합추정량의 평균제곱오차를 근사적으로 설명할 수 있는 분산함수 (variance function)를 활용하는 방안을 향후 검토하고자 한다.

<표 4.9> 제조업 및 도소매업 표본크기의존 복합추정 결과

중 분 류	사업체수		평균임금	소 분 류	사업체수		평균임금 (직접추정)	가중회귀 (간접추정)	복합추정(1)		복합추정(2)	
	모집단	표 본			모집단	표 본			가중값	추정값	가중값	추정값
15	5045	200	1,736,155	151	1954	72	1,351,370	1,435,850	0.929	1,357,328	0.896	1,360,147
				152	115	8	2,058,807	2,080,241	1	2,058,807	1	2,058,807
				153	756	30	2,102,012	1,820,353	1	2,102,012	1	2,102,012
				154	1838	72	1,724,251	1,806,786	0.988	1,725,230	0.982	1,725,715
				155	382	18	2,252,257	2,040,201	1	2,252,257	1	2,252,257
17	7601	152	1,471,344	171	998	44	1,563,533	1,530,023	1	1,563,533	1	1,563,533
				172	1589	37	1,413,727	1,418,008	1	1,413,727	1	1,413,727
				173	1387	17	1,470,850	1,516,003	0.613	1,488,328	0.48	1,494,336
				174	1520	34	1,468,085	1,487,342	1	1,468,085	1	1,468,085
				179	2107	20	1,366,590	1,367,902	0.475	1,367,279	0.327	1,367,473
24	3742	105	2,365,373	241	1411	48	2,599,119	2,469,603	1	2,599,119	1	2,599,119
				242	601	19	2,224,997	2,312,695	1	2,224,997	1	2,224,997
				243	1641	32	1,884,409	1,999,891	0.695	1,919,636	0.579	1,932,987
				244	89	6	3,128,951	3,076,764	1	3,128,951	1	3,128,951
32	4787	94	1,849,678	321	321	50	1,847,879	1,823,200	1	1,847,879	1	1,847,879
				322	322	23	1,918,487	1,881,141	1	1,918,487	1	1,918,487
				323	323	21	1,791,318	1,890,498	1	1,791,318	1	1,791,318
50	4629	147	1,754,904	501	1799	65	2,143,221	2,041,276	1	2,143,221	1	2,143,221
				502	831	33	1,668,215	1,642,586	1	1,668,215	1	1,668,215
				503	49	3	1,913,215	1,956,734	1	1,913,215	1	1,913,215
				504	1950	46	1,238,953	1,406,069	0.743	1,281,929	0.64	1,299,075
51	26472	232	2,008,664	511	1808	33	2,454,652	2,124,645	1	2,454,652	1	2,454,652
				512	484	7	1,723,209	1,963,808	1	1,723,209	1	1,723,209
				513	4740	42	1,938,054	1,865,592	1	1,938,054	1	1,938,054
				514	6120	62	1,707,198	1,839,264	1	1,707,198	1	1,707,198
				515	1899	18	2,177,728	2,158,733	1	2,177,728	1	2,177,728
				516	1229	7	2,239,630	2,218,457	0.65	2,232,217	0.524	2,229,550
				517	2904	14	1,887,911	1,961,074	0.55	1,920,828	0.408	1,931,225
				518	5567	41	2,098,383	2,204,086	0.84	2,115,258	0.77	2,122,657
				519	1721	8	2,408,186	2,250,954	0.53	2,334,351	0.386	2,311,691
52	12633	150	1,601,072	521	2551	53	1,704,224	1,699,208	1	1,704,224	1	1,704,224
				522	863	10	1,269,149	1,320,739	0.976	1,270,392	0.964	1,271,002
				523	1532	17	1,511,742	1,544,809	0.935	1,513,906	0.903	1,514,935
				524	1194	13	1,585,718	1,498,398	0.917	1,578,468	0.878	1,575,071
				525	2197	16	1,496,402	1,499,396	0.613	1,497,559	0.48	1,497,958
				526	2730	34	1,526,678	1,570,276	1	1,526,678	1	1,526,678
				528	1566	7	1,720,150	1,635,646	0.376	1,667,459	0.231	1,655,165

4.6 산업소분류 평균임금 추정을 위한 EBLUP 추정

직접추정량과 합성추정량을 결합하는 4.5절에 제시된 복합추정량 대신에 모형기반(model based) 소지역 추정기법으로 소지역 요인을 랜덤효과(random effect)로 모형에 반영한 혼합선형모형(mixed linear model)을 이용하는 방법을 고려할 수 있다(Rao, 2003).

임금구조기본통계조사의 경우 각 근로자에 대한 설명변수가 조사되고 있기 때문에 소지역 추정을 위한 혼합선형모형 중 다음과 같은 단위수준(unit level) 소지역 모형을 활용한 EBLUP(Empirical Best Linear Unbiased Predictor) 추정이 가능하다. EBLUP 추정을 위한 혼합선형모형은 다음과 같이 요약될 수 있다.

특정 산업 중분류내의 근로자 중 i 소분류내 j 번째 근로자의 총임금을 y_{ij} 라고 하자.

$$y_{ij} = x_{ij}^T \beta + v_i + e_{ij} \quad j=1, \dots, n_i, \quad i=1, \dots, m \quad (4.5)$$

여기서 x_{ij} 는 4.3절에서 선택된 설명변수로 구성된 공분량, v_i 는 i 산업소분류의 특성을 반영하는 랜덤효과, e_{ij} 는 오차항을 나타낸다.

본 연구에서 추정대상은 산업 소분류별 평균임금에 해당하는 \bar{Y}_i 이고 이에 대한 추정량은 다음의 \widehat{Y}_i^H 와 같이 표현되고, 이를 BLUP(Best Linear Unbiased Predictor)라고 한다(Henderson, 1950). 추출률 n_i/N_i 은 무시할 수 있다고 가정한 것이다.

$$\widehat{Y}_i^H = \bar{X}_i^T \beta + \gamma_i (\bar{y}_{ia} - \bar{x}_{ia}^T \beta) \quad (4.6)$$

여기서 $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / a_i)$, $\bar{y}_{ia} = \sum_j a_{ij} y_{ij} / a_i$, $\bar{x}_{ia} = \sum_j a_{ij} x_{ij} / a_i$ 이고, β 는 β 에 대한 BLUE(best linear unbiased estimator) 이다.

한편 BLUP 추정량 (4.6)을 다시 표현하면 다음과 같이 조사회귀(survey regression) 추정량 $\bar{y}_{ia} + (\bar{X}_i - \bar{x}_{ia})^T \beta$ 과 회귀합성추정량 $\bar{X}_i^T \beta$ 의 가중평균으로 나타낼 수 있다. 따라서 BLUP은 (4.1)의 복합추정량과 유사한 형태의 추정량임을 알 수 있다.

$$\widehat{Y}_i^H = \gamma_i [\bar{y}_{ia} + (\bar{X}_i - \bar{x}_{ia})^T \beta] + (1 - \gamma_i) \bar{X}_i^T \beta \quad (4.7)$$

식(4.6)의 BLUP에서 가중값에 해당하는 $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / a_i)$ 를 적절한 방법으로 구한 추정값 $\widehat{\gamma}_i$ 로 대체한 추정량을 EBLUP 추정량이라고 한다. 여기서 \bar{X}_i 는 모집단에서 설명변수 x_{ij} 들의 평균에 해당하는 것으로 임금구조통계조사의 경우 산업소분류별 모집단에서 설명변수의 평균에 해당하는 \bar{X}_i 에 대한 정보를 현실적으로 얻을 수 없다. 따라서 불가피하게 \bar{X}_i 대신 표본에서 설명변수의 가중평균에 해당하는 \bar{x}_{ia} 를 사용하였다. 이에 따라 본 연구에서는 다음과 같이 변형된 EBLUP 추정량을 사용한 것이다.

$$\bar{Y}_{(EBLUP),i} = \hat{\gamma}_i \bar{y}_{ia} + (1 - \hat{\gamma}_i) \bar{x}_{ia}^T \beta \quad (4.8)$$

제시된 EBLUP 추정량 $\bar{Y}_{(EBLUP),i}$ 은 (4.1)의 복합추정량과 거의 동일한 추정결과를 제공한다. (4.8)에서는 가중값 $\hat{\gamma}_i$ 이 혼합선형모형의 적합과정에서 반복과정을 통해 우도함수를 최대화시키는 값으로 추정되고 있기 때문에 결과적으로 (4.8)의 EBLUP 추정량은 (4.1)에서 최적의 가중값 (4.2)를 적용한 경우와 거의 동일한 추정결과를 산출하게 된다.

각 산업소분류의 평균임금에 대한 EBLUP 추정값은 SAS의 Mixed Procedure를 통해 얻어진 각 근로자에 대한 예측결과를 활용하여 산출될 수 있다. 제조업 및 도소매업의 일부 산업소분류에 대한 이런 과정을 통해 얻어진 EBLUP 추정결과는 <표 4.10>과 같다.

EBLUP에서 사용된 (4.5)의 모형은 근로자를 관측값으로 하기 때문에 실제 혼합선형모형을 적합시키는 데 있어서 각 산업소분류의 근로자수가 표본크기로 반영된다. 본 연구에서 분석대상으로 하는 제조업 및 도소매업의 일부 산업소분류에서는 <표 4.10>에 나타난 것과 같이 근로자수를 기준으로하면 각 소분류별 표본크기가 충분히 크기 때문에 $\hat{\gamma}_i$ 는 1에 가까워지고 따라서 EBLUP 추정결과는 직접추정값 \bar{y}_{ia} 와 매우 유사하다는 것을 알 수 있다. 이는 제시된 EBLUP 추정방법이 매우 효과적으로 적용되고 있다는 것을 입증하고 있다. 특히 소지역 추정기법의 도입이 필요한 표본 근로자수가 충분히 크지 않은 산업소분류의 경우 이런 EBLUP 추정법을 통해 매우 효율적인 산업소분류별 평균임금에 대한 추정값을 얻을 수 있을 것으로 예상된다.

한편 본 연구에선 제시된 EBLUP 추정법은 기존의 방법을 변형한 것으로 이와 관련된 정확도를 설명하기 위한 평균오차제곱(MSE)의 산출을 위해서는 차후 이와 관련된 이론에 대한 추가적인 연구가 필요하다.

<표 4.10> 제조업 및 도소매업 EBLUP 추정결과

중분류	가중평균	소분류	사업체 표본수	근로자 표본수	가중평균	가중회귀	EBLUP추정
15	1,736,155	151	72	4205	1,351,370	1,435,850	1,351,898
		152	8	1081	2,058,807	2,080,241	2,059,416
		153	30	985	2,102,012	1,820,353	2,100,349
		154	72	5829	1,724,251	1,806,786	1,724,638
		155	18	2032	2,252,257	2,040,201	2,250,950
17	1,471,344	171	44	5223	1,563,533	1,530,023	1,560,441
		172	37	1508	1,413,727	1,418,008	1,413,788
		173	17	339	1,470,850	1,516,003	1,480,308
		174	34	2405	1,468,085	1,487,342	1,469,503
		179	20	427	1,366,590	1,367,902	1,365,157
24	2,365,373	241	48	3891	2,599,119	2,469,603	2,598,291
		242	19	3235	2,224,997	2,312,695	2,226,146
		243	32	2894	1,884,409	1,999,891	1,884,859
		244	6	1960	3,128,951	3,076,764	3,128,070
32	1,849,678	321	50	6732	1,847,879	1,823,200	1,847,591
		322	23	2078	1,918,487	1,881,141	1,917,963
		323	21	3284	1,791,318	1,890,498	1,792,559
50	1,754,904	501	65	1192	2,143,221	2,041,276	2,138,846
		502	33	921	1,668,215	1,642,586	1,667,623
		503	3	46	1,913,215	1,956,734	1,927,098
		504	46	537	1,238,953	1,406,069	1,245,145
51	2,008,664	511	33	1151	2,454,652	2,124,645	2,448,409
		512	7	175	1,723,209	1,963,808	1,739,573
		513	42	1534	1,938,054	1,865,592	1,937,090
		514	62	2380	1,707,198	1,839,264	1,708,651
		515	18	637	2,177,728	2,158,733	2,177,858
		516	7	456	2,239,630	2,218,457	2,236,094
		517	14	458	1,887,911	1,961,074	1,890,949
		518	41	2793	2,098,383	2,204,086	2,100,204
		519	8	573	2,408,186	2,250,954	2,398,961
52	1,601,072	521	53	7490	1,704,224	1,699,208	1,703,954
		522	10	140	1,269,149	1,320,739	1,275,452
		523	17	206	1,511,742	1,544,809	1,514,605
		524	13	371	1,585,718	1,498,398	1,579,625
		525	16	192	1,496,402	1,499,396	1,497,008
		526	34	493	1,526,678	1,570,276	1,528,764
		528	7	931	1,720,150	1,635,646	1,712,454

4.7 산업소분류 평균임금 추정결과 비교 및 향후 과제

본 연구에서는 산업 소분류별 평균임금을 추정하기 위한 소지역 추정기법으로 4.5절에서 직접추정량과 회귀모형을 이용한 간접추정량을 가중평균한 형태의 복합추정법을 제시하였고, 4.6절에서는 모형기반 소지역 추정기법에 해당하는 EBLUP 추정법을 제시하였다. 그 결과를 종합하여 정리하면 <표 4.11>과 같다.

<표 4.11> 산업 소분류 평균임금 추정결과(종합)

중분류	평균임금	소분류	직접추정	합성추정	복합추정(1)	복합추정(2)	EBLUP추정
15	1,736,155	151	1,351,370	1,435,850	1,357,328	1,360,147	1,351,898
		152	2,058,807	2,080,241	2,058,807	2,058,807	2,059,416
		153	2,102,012	1,820,353	2,102,012	2,102,012	2,100,349
		154	1,724,251	1,806,786	1,725,230	1,725,715	1,724,638
		155	2,252,257	2,040,201	2,252,257	2,252,257	2,250,950
17	1,471,344	171	1,563,533	1,530,023	1,563,533	1,563,533	1,560,441
		172	1,413,727	1,418,008	1,413,727	1,413,727	1,413,788
		173	1,470,850	1,516,003	1,488,328	1,494,336	1,480,308
		174	1,468,085	1,487,342	1,468,085	1,468,085	1,469,503
		179	1,366,590	1,367,902	1,367,279	1,367,473	1,365,157
24	2,365,373	241	2,599,119	2,469,603	2,599,119	2,599,119	2,598,291
		242	2,224,997	2,312,695	2,224,997	2,224,997	2,226,146
		243	1,884,409	1,999,891	1,919,636	1,932,987	1,884,859
		244	3,128,951	3,076,764	3,128,951	3,128,951	3,128,070
32	1,849,678	321	1,847,879	1,823,200	1,847,879	1,847,879	1,847,591
		322	1,918,487	1,881,141	1,918,487	1,918,487	1,917,963
		323	1,791,318	1,890,498	1,791,318	1,791,318	1,792,559
50	1,754,904	501	2,143,221	2,041,276	2,143,221	2,143,221	2,138,846
		502	1,668,215	1,642,586	1,668,215	1,668,215	1,667,623
		503	1,913,215	1,956,734	1,913,215	1,913,215	1,927,098
		504	1,238,953	1,406,069	1,281,929	1,299,075	1,245,145
51	2,008,664	511	2,454,652	2,124,645	2,454,652	2,454,652	2,448,409
		512	1,723,209	1,963,808	1,723,209	1,723,209	1,739,573
		513	1,938,054	1,865,592	1,938,054	1,938,054	1,937,090
		514	1,707,198	1,839,264	1,707,198	1,707,198	1,708,651
		515	2,177,728	2,158,733	2,177,728	2,177,728	2,177,858
		516	2,239,630	2,218,457	2,232,217	2,229,550	2,236,094
		517	1,887,911	1,961,074	1,920,828	1,931,225	1,890,949
		518	2,098,383	2,204,086	2,115,258	2,122,657	2,100,204
		519	2,408,186	2,250,954	2,334,351	2,311,691	2,398,961
52	1,601,072	521	1,704,224	1,699,208	1,704,224	1,704,224	1,703,954
		522	1,269,149	1,320,739	1,270,392	1,271,002	1,275,452
		523	1,511,742	1,544,809	1,513,906	1,514,935	1,514,605
		524	1,585,718	1,498,398	1,578,468	1,575,071	1,579,625
		525	1,496,402	1,499,396	1,497,559	1,497,958	1,497,008
		526	1,526,678	1,570,276	1,526,678	1,526,678	1,528,764
		528	1,720,150	1,635,646	1,667,459	1,655,165	1,712,454

<표 4.11>에서 직접추정은 조사 자료에서 설계가중값을 반영해 직접 산출한 소분류별 평균임금 추정결과, 합성추정은 4.4절에 제시된 설명변수를 반영한 선형모형을 통해 얻어진 평균임금 추정결과, 복합추정[1]과 복합추정[2]는 4.5절에 제시된 표본크기의존 복합추정결과, 그리고 EBLUP 추정은 4.6절에 제시된 혼합선형모형을 기반으로 얻어진 추정결과를 의미한다. 소지역 추정기법을 적용한 복합추정결과들과 EBLUP추정결과를 비교해보면 전반적으로 큰 차이를 보이고 있지 않는 것으로 보아 이들 추정기법들이 상당수준 신뢰성 있는 추정결과를 산출하고 있다는 것으로 해석할 수 있다. 물론 이와 관련해서는 각 추정법의 정확도를 명확히 측정할 수 있는 평균제곱오차를 산출해 비교하는 작업이 차후 수행되어야 한다.

한편 본 연구에서는 산업 소분류별 평균임금 산출을 위해 적절한 소지역 추정기법을 제시하고, 이에 따라 제한적으로 제조업 및 도소매업의 일부 소분류에 대해 제시한 방법에 의한 평균임금 추정값을 구하였다. 하지만 이들 추정기법들을 본격적으로 공식적인 노동통계 산출에 적용하기 위해서는 주어진 방법들의 효율성을 검토하는 작업이 필요하다. 따라서 이들 추정기법의 효율성 검토를 위해서는 각 추정량의 정확도를 설명할 수 있는 평균제곱오차(MSE)를 산출하기 방안에 대한 이론적인 연구가 향후 수행되어야 한다. 아울러 이런 과정을 통해 로그변환 등과 같은 기법을 추가적으로 도입함으로써 제시된 기법의 효율성을 보다 향상시킬 수 있는 방안을 모색하는 작업이 필요하다고 판단된다.

제 5 장 직종세분류 평균임금 산출을 위한 소지역 추정기법

5.1 고용구조통계조사 개요

(1) 산업·직종별 고용구조조사 개요

고용구조조사는 조사대상을 취업자의 세부 산업·직종별 고용구조를 파악하는 것이므로 15세 이상 경제활동인구 가운데 취업자를 조사대상으로 한다. 조사를 위한 기준 산업분류는 한국표준산업분류(소분류 단위)로 하였으며, 직업분류는 고용직업분류를 중심으로 노동시장의 상황에 맞게 재편된 중앙고용정보원 직업분류를 이용하고 있다.

조사방식은 조사항목을 조사대상자에게 질문하고 조사대상자의 답변을 조사원이 기입하는 면접 단계식을 기본으로 하며, 필요시 전화조사를 병행하여 실시하였다. 산업 및 직업관련 부호화는 일차적으로 직업상담원에 의해 분류·코딩되었으며, 최종적으로 전 표본에 대하여 중앙고용정보원의 직업 연구원에 의해 재 코딩되었다.

(2) 설문구성

산업·직종별 고용구조조사는 취업자를 대상으로 산업별, 직업별 고용인원을 파악하는 조사라는 점에서 미국의 OES조사와 가장 유사한 성격을 띠고 있다. 그러나 미국의 OES는 사업체를 대상으로 이루어지는 사업체 조사로써 취업자 중 임금근로자에 대한 고용인원만 파악된다는 한계를 가진다. 우리나라는 취업자 중에서 비임금근로자의 비중이 상대적으로 높은 점을 감안하면 조사시 임금근로자 뿐만 아니라 비임금근로자에 대한 고용인원 파악이 요구된다. 따라서 임금/비임금의 모든 종사상 지위에 대하여 조사가 이루어지는 ‘경제활동인구조사’ 설문지를 기본으로 하되, ‘경제활동인구조사’에서 조사되지 않는 항목인 임금은 미국의 OES를 참고하여 작성되었다.

산업·직업별 고용구조조사 설문은 크게 주요항목과 직업과 관련한 보조항목으로 구성되었다. 주요항목은 194개의 산업 소분류와 419개의 직업 세분류를 정확히 파악하여 분류할 수 있도록 단계적으로 이루어졌다. 산업 소분류를 파악하기 위하여 우선 조사대상자가 소속된 기업체나 사업체의 명칭에 대하여 질문하고, 소속된 사업체의 산출물이나 서비스 내용에 대한 질문을 추가함으로써 분류가 쉽도록 하였다.

직업 세분류는 우선 직위(직급), 부서명 및 현장에서 불리는 직업 명칭에 대하여 파악한 후, 직업분류를 위해서 좀더 상세한 정보를 얻기 위하여 조사대상자의 구체적인 직무내용에 대한 항목을 질문하였다.

보조항목은 크게 인적사항 및 노동시장 정보로 구성되며, 이 두 정보는 직업 분류를 용이하게 하

며, 향후 분류의 신뢰성 여부 판단에 중요한 영향을 미치는 항목들이다. 인적사항은 6개 항목 즉 성명, 성별, 생년월일, 혼인상태, 가구주와의 관계, 교육내용(학력, 전공, 이수여부)로 이루어진다. 노동시장 정보 역시 6개 항목 즉 종사상 지위(임금/비임금), 근속년수, 근로시간, 자격증, 임금, 사업체 종사자 주로 구성된다.

(3) 표본설계

고용구조조사는 산업 및 직업별 고용자수와 임금수준을 파악하기 위한 것으로 가구를 대상으로 조사하고 있다. 우리나라는 고용자, 자영업자 등 비임금근로자 비율이 다른 나라에 비해 높아서 임금근로자만이 조사대상으로 선정되는 사업체조사방식에는 한계가 있다. 일반적으로 가구조사방식은 사업체조사보다 높은 거부율 등 몇 가지 문제점을 가지고 있으나 노동시장의 대표성을 감안하여 가구조사방식을 선택했다.

<표 5-1>노동시장에 따른 조사방식 비료

	OES (occupational employment statistics, 미국)	LFS(labor force survey, 캐나다)	산업.직업별 고용구조조사 (WIC-OES, 한국)
조사대상	5인 이상 사업체	일반가구	취업자 가구
조사방법	우편조사	가구방문, 전화조사	가구방문
주요항목	산업, 직업(700), 임금	경제활동상태 및 산업, 직업(520), 임금	산업, 직업(419), 임금
특징	고용인원파악에 중점	경제활동인구조사와 유사	고용인원, 임금수준
임금근로자 비율(2001년)	93%	85%	62%

출처 : 박천수 외 3인(2001)

고용구조조사의 조사모집단은 1995년도 인구주택 총조사 결과를 사용하였고, 10% 표본조사구 중에서 섬지역 조사구(제주도는 포함)와 시설단위 조사구를 제외한 22,029개의 보통 조사구를 추출단위로 하는 표본틀을 구성하였다. 표본설계 당시에 결정된 각 지역별 표본조사구 및 표본가구 수는 <표 5-2>와 같다.

<표 5-2> 지역별 표본규모

	조사구수	표본조사구	표본가구수
서울	5,119	567	11,340
부산	1,843	210	4,200
대구	1,216	139	2,780
대전	596	69	1,380
인천	1,116	127	2,540
광주	608	70	1,400
울산	456	53	1,060
경기	3,652	419	8,380
강원	742	84	1,680
충북	699	79	1,580
충남	874	100	2,000
전북	967	109	2,180
전남	998	116	2,320
경북	1,375	161	3,220
경남	1,425	167	3,340
제주	263	30	600
전체	22,029	2,500	50,000

출처 : 박천수 외 3인(2001)

이 조사에 사용되는 표본 조사구는 7대 대도시와 각 도의 동부와 읍·면부의 총 25개 지역으로 층화한 후 조사구 명부서 조사구 내의 가구수의 크기에 비례하는 확률비례계통추출법(probability proportional to sampling)을 적용하여 표본조사구를 추출하였다. 표본가구는 추출된 2,500개 표본조사구에서 각각 20가구씩을 계통추출법(systematic random sampling)에 의해 추출하였다. 결국 조사대상가구는 지역, 주택특성 산업구성비 등이 고려된 분류지료를 감안하여 조사구내의 가구수 규모의 크기에 비례하는 확률비례계통추출법을 적용한 2단 집락추출법(Two-stage cluster sampling)에 의해 추출된 것이다.

(4) 가중값 부여

고용구조조사의 가중값은 사후층화를 통한 가중값 부여방식을 적용하여 각 산업 및 직업별 종사자 인원에 대한 모수를 추정한다. 향후 직업별 고용전망을 위해서는 전체 경제활동인구와의 연결이 필요하다. 「경제활동인구조사」의 결과를 기초로 성별, 지역 및 산업대분류 통계수치를 이용하여 가중값을 산출하였다. 첫째로 「경제활동인구조사」의 성별×지역별 가중값을 적용하여 산업대분류별 취업자수를 구하고 이 결과를 「경제활동인구조사」의 산업대분류별 취업자수를 나누어 2차 가중값을 구한다. 이러한 과정을 반복하여 「경제활동인구조사」의 결과와 모수를 알 수 있는 전문가 직종과의 오차를 최소화하는 가중값을 최종적으로 결정하였다.

(5) WIC-OES 직업분류의 구성체계

WIC-OES 직업분류는 “산업·직업별 고용조사”와 “청년패널 구축을 위한 사전조사”에 활용되기 위하여 작성되었다. WIC-OES 분류는 기존 직업분류들을 총체적으로 분석하고, 직무 및 고용 단위 구분이 용이한 최근 개발된 「고용직업분류」를 수정·보완한 것으로 기존 분류들이 갖는 단점들을 개선한 것이다.

WIC-OES 직업분류체계에서 분류원칙은 직무유형과 직능수준을 기준으로 하되 직업별 종사자 규모도 고려되었다. 대분류는 고용직업분류의 분류틀을 준용하여 산업적 특성이 많이 반영되었다. 즉, 직무유형(Skill Type)을 기준으로 0~9단위까지 10개의 직업대분류가 구성되었고, 그 이하 중, 소, 세분류는 직무의 유사성, 직능수준, 직업의 이동성이 고려되었다. 또한 임금과 직업전망 등을 고려한 직업정보로서의 가치가 중시되었다. WIC-OES 직업분류의 특징은 직업대분류가 직무유형(Skill Type)을 기준으로 분류되었기에 산업과의 연계가 가능하고, 일반인들의 직업인식에도 잘 부합된다는 점이다.

<표 5-3> WIC-OES 직업분류체계의 분류원칙

	대분류	중·소·세분류	기타
분류원칙	직무유형	직무유형, 직능수준, 직업의 이동성, 종사자 규모	직업정보의 가치

출처 : 박천수 외 3인(2001)

WIC-OES 직업분류는 대분류, 중분류, 소분류 및 세분류의 4단계로 구성되어 있으며 대분류는 10개, 중분류는 38개, 소분류는 146개, 세분류는 419개이다. 각 분류 단계별 항목 수는 다음 <표 5-5>와 같다.

<표 5-5> WIC-OES 직업분류의 분류단계별 항목 수

대분류	중분류	소분류	세분류
0. 관리직	10	19	35
1. 사업, 금융 및 사무관련직	4	13	39
2. 과학, 공학 및 정보시스템 관련직	4	16	44
3. 교육 및 법률 관련직	2	6	16
4. 보건의료, 사회서비스 및 종교 관련직	3	14	32
5. 문화, 예술 및 스포츠 관련직	2	11	38
6. 판매 및 서비스 관련직	2	13	51
7. 기능 및 운송 관련직	5	30	89
8. 제조 및 전기, 가스, 수도설비 관련직	4	17	63
9. 농림어업 관련직	2	4	12
합계	38	143	419

출처 : 박천수 외 3인(2001)

5.2 고용구조조사 결과를 이용한 직종세분류별 평균임금 추정 연구방향

(1) 분석방향

본 연구에서 주요 자료로 이용하고 있는 “2002년 임금구조기본통계조사”는 근로자가 소속된 사업체의 산업분류는 소분류까지 조사하고 있고, 직종에 대해서는 소분류까지 조사하고 있다. 이런 이유로 현재 상태에서 임금구조기본통계조사 결과만을 이용해서 직종세분류별 근로자 수와 평균임금을 추정하는 것은 불가능하다. “2003년 임금구조기본통계조사”부터는 시험적으로 근로자의 직종세분류까지 조사하고 있어서 2004년 연구에서는 사업체 조사결과에 대해서 직종세분류별 평균임금 추정이 가능할 것이다.

고용구조조사는 가구를 대상으로 하는 조사로 기본적으로 사업체를 대상으로 하는 조사와 차이가 나지만, 사업체조사 결과를 보완하고 점검할 수 있을 것이다. 특히 현재 고용구조조사는 각 취업자의 직종세분류까지 조사하고 있어서 취업자의 직종세분류별 평균임금 작성 가능성을 검토하는데 활용될 수 있다. 본 연구에서 고용구조조사를 분석하는 목적은 다음과 같이 요약할 수 있다.

- 소지역통계기법을 이용한 직종세분류별 평균 임금추정의 가능성 검토
- 사업체조사 결과의 소지역 통계 작성 결과와의 비교 검토
- 고용구조조사 결과를 사업체 조사와 결합해서 이용할 수 있는 방안 검토

이와 같은 분석 목적을 달성하기 위해서 다음과 같이 분석방향을 정하였다.

첫째, 산업·직업별 고용구조 조사 결과 중 근로자 5인 이상인 사업체에 종사하는 임금근로자의 정규직만을 분석 대상으로 한다. 현재 임금구조기본통계조사는 상용근로자 5인 이상이 종사하는 사업체를 대상으로 상용근로자만을 대상으로 조사하고 있다. 따라서 분석된 두 결과를 비교하기 위해서는 분석대상이 유사해야 할 것이다. 물론 고용구조조사는 가구원의 응답에 기초한 조사이고, 임금구조기본통계조사는 사업체의 임금대장에 기초하여 조사하기 때문에 조사방법에 따라서 평균임금에 차이가 있을 것으로 예상된다.

둘째, 가능하면 임금구조기본통계조사 결과를 분석할 때 사용된 통계모형과 같은 방법론을 적용해서 분석한다. 이와 같이 같은 방법론을 적용함으로써 직종세분류별 평균 임금추정의 가능성을 검토할 수 있고, 향후 사업체조사 결과의 소지역 통계 작성 결과와의 비교할 수 있을 것이다.

셋째, 고용구조조사 분석결과는 시험적인 연구로 임금구조기본통계조사 결과의 분석결과와 비교하기 위해서 산업중분류 15, 17, 24, 31, 32, 33 등을 대상으로 한다. 또한 직종세분별 분석은 “운전 및 운송 관련직 전체”, “환경, 인쇄, 목재, 가구, 공예 및 생산단순직 전체”를 대상으로 한다.

(2) 분석 데이터 현황

다음 <표 5-6>은 2002년 고용구조조사의 전체 조사결과와 분석대상에 대한 지역별 현황을 정리한 것이다. 본 연구에서 사용되는 분석대상은 전체 조사결과 중에서 약 39%이다.

<표 5-6> 지역별 전체 조사결과 및 분석대상 현황

시도 구분	전체 조사결과			분석대상		
	조사구수	가구수	조사대상	조사구수	가구수	조사대상
서울	533	10666	15001	533	5830	7061
부산	197	3776	5422	195	1848	2204
대구	131	2704	3865	129	1149	1349
인천	71	1443	1956	71	683	794
광주	129	2708	3391	128	1504	1685
대전	72	1641	2242	72	888	1043
울산	51	1019	1483	51	635	796
경기	454	8463	11767	451	4818	5818
강원	88	1687	2498	78	527	617
충북	82	1671	2531	77	667	776
충남	104	2205	3518	94	809	919
전북	110	2348	3511	99	833	957
전남	111	2437	3775	86	591	661
경북	153	3093	4778	125	984	1134
경남	164	2883	4694	153	1282	1603
제주	29	572	928	27	192	222
합계	2479	49316	71360	2369	23240	27639

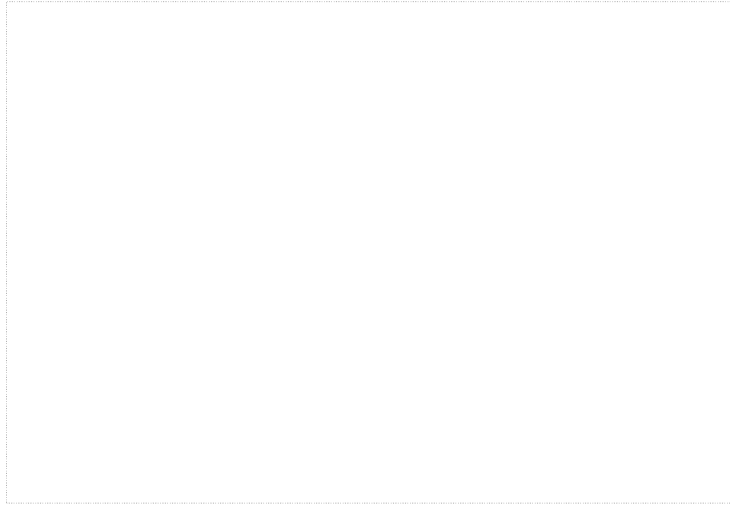
다음 <표 5-7>은 분석대상 데이터에 대한 산업대분류별, 사업체 규모별 응답자 현황을 정리한 것이다.

<표 5-7> 산업대분류별, 사업체 규모별 응답자 현황

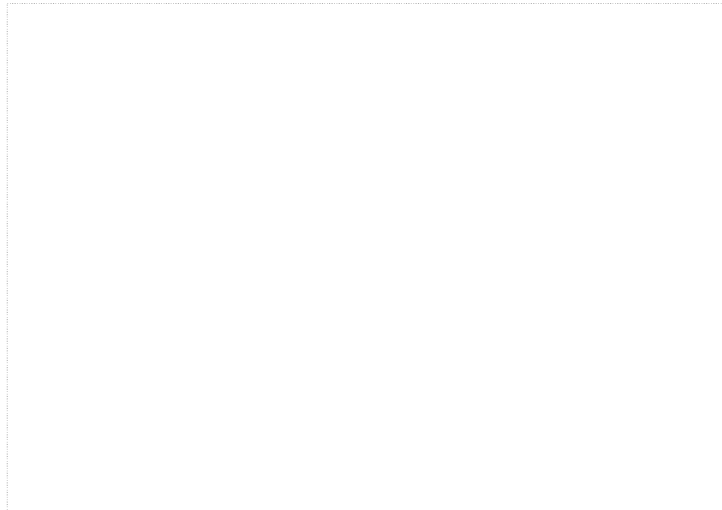
빈도	5-9인	10-29인	30-99인	100-299인	300-499인	500인-	총합
C	4	6	12	6	2	6	36
D	1026	1689	1969	1221	467	1751	8123
E	21	31	67	63	7	60	249
F	573	675	374	183	36	88	1929
G	721	734	487	253	102	208	2505
H	348	208	93	51	17	41	758
I	169	358	540	467	116	208	1858
J	72	106	179	138	39	72	606
K	153	607	590	228	78	183	1839
L	168	220	75	15	7	5	490
M	445	613	498	227	67	174	2024
N	166	411	374	399	189	411	1950
O	385	653	1541	270	64	152	3065
P	193	180	218	127	58	188	964
Q	48	82	113	63	17	47	370
R	273	279	161	76	18	26	833
총합	4765	6852	7291	3787	1284	3620	27599

(3) 분석 준비 작업

임금구조기본통계조사의 결과 비교를 위해서 1-4인 규모의 사업체에 종사하는 응답자는 분석 대상에서 제외하였고, 농.임.어업 등의 1차 산업은 근로자 수가 적어서 제외하였다. 한편, 통계 모형에 사용될 변수들은 임금구조기본통계조사에서 사용된 분류 방식으로 정리하였다. 예를 들어 학력(final_ed)은 무학(1)과 초등학교(2) 졸업을 묶었고, 석사과정(7)+박사과정(8)을 묶었다. 통계분석을 위해서는 분석 전에 반드시 데이터 cleaning 절차가 필요하다. 근로자의 임금은 우선 이상값을 검출해서 제거하고, 임금의 로그변환한 값을 통계모형의 종속변수로 이용하였다. 다음 <그림 5-1>과 <그림 5-2>는 각각 임금 원자료와 로그변환된 값에 대한 히스토그램을 나타낸 것이다. 로그 변환 전의 원자료를 살펴보면 왼쪽으로 치우친 감마분포(gamma distribution) 형태를 나타내지만, 로그 변환 후의 값은 대략 좌우대칭인 형태를 보인다.



<그림 5-1> 로그변환 전의 임금자료의 히스토그램



<그림 5-2> 로그변환 후의 임금자료의 히스토그램

임금의 로그변환값에 대한 통계모형 적합은 일반화선형모형 접근법을 이용한다. 통계모형에 사용된 독립변수들은 임금구조기본통계조사 분석에서 사용된 것으로 한다.

▪ 일반화선형모형 접근법(EBLUP 방법)

- 종속변수 : 로그변환된 임금자료 ($\ln(wage)$)
- 설명변수 : 산업, 직종, 학력, 성별, 나이, 경력년수, 사업체 규모, 근로시간, 결혼여부, 지역(시도 구분), 임금형태(월급, 연봉제) 등

적용되는 일반화선형모형에서 직종분류는 3자리까지는 fixed effect로 간주하고, 4자리는 random effect로 처리한다. 산업분류는 2자리까지는 fixed effect로 간주하고, 3자리는 random effect로 처리한다. 이러한 방법을 적용하면 추정되는 모수의 수가 많이 줄어들어 회귀계수 추정값이 좀더 안정적이다.

본 연구의 분석결과는 시험적인 연구로 임금구조기본통계조사 결과의 분석결과와 비교하기 위해서 다음에 제시된 산업소분류와 직종세분류에 대해서 분석결과를 수록한다.

▪ 산업소분류별 분석 :

산업중분류 15, 17, 24, 31, 32, 33 (사업체 조사 결과 분석과 비교 목적)

▪ 직종세분류별 분석

- 운전 및 운송 관련직 전체 :

- 환경, 인쇄, 목재, 가구, 공예 및 생산단순직 전체 :

본 연구에서는 설명변수로 사용된 변수 중에서 결측치가 있으면 이를 분석에서 제외하였다. 향후 연구에서는 분석에 앞서서 결측치 현황을 파악하여 결측치에 대해서는 imputation 기법을 적용해서 설명변수 데이터를 정리해야 할 것이다. 또한 몇 년간의 연구결과가 축적되면 소지역통계 기법 중 시계열 접근법을 적용해서 좀더 안정적인 추정이 가능할 것이다.

5.3 소지역 추정을 위한 선형모형의 적합

최종적으로 분석에 사용된 자료는 모두 22,021명의 조사결과이다. 최종적인 분석에 사용된 자료는 임금구조기본통계조사의 조사대상이 아닌 'N. 공공행정'과 이상값을 제외한 것이다. 통계모형에 사용된 변수 중에서 나이 변수를 제외하면 임금에 모두 유의한 영향을 미치는 것으로 나타났다. 특히, 성별, 경력년수, 최종학력, 결혼여부, 임금형태 등이 중요한 변수로 나타났다.

<표 5-8> 적합된 일반화선형모형에서 유의성 정도

Effect	DF	Num	Den	F Value	Pr > F
		DF			
matro	15	21E3		28.36	<.0001
sex	1	21E3		1236.93	<.0001
age	1	21E3		0.00	0.9452
career_y	1	21E3		1595.21	<.0001
final_ed1	5	21E3		234.76	<.0001
kumo	5	21E3		65.83	<.0001
worktime	1	21E3		27.86	<.0001
job1	116	21E3		3.53	<.0001
ind1	55	21E3		1.73	0.0006
mar	1	21E3		554.64	<.0001
wagetype	1	21E3		1823.82	<.0001

다음 <표 5-9>는 산업대분류별 평균임금에 대해서 직접추정량과 소지역추정기법을 적용하여 구한 평균임금을 정리한 것이다. 전체적으로 평균임금이 적은 산업분류에 대한 소지역추정기법 적용 결과는 다소 커지는 경향이 있고, 평균임금이 상대적으로 큰 경우에는 소지역추정기법을 적용한 결과가 다소 작아지는 경향을 나타내고 있다.

<표 5-9> 산업대분류별 평균임금에 대한 직접추정량과 소지역추정기법의 비교

산업대분류	표본크기	직접추정량		소지역추정기법
		평균임금	CV(%)	평균임금
C	33	177.82	7.66%	176.75
D	7409	172.02	1.12%	171.86
E	235	235.24	2.85%	236.65
F	1513	195.26	1.50%	197.20
G	2010	162.70	1.25%	161.93
H	491	133.39	2.13%	134.68
I	1666	166.71	1.38%	167.58
J	532	210.90	2.16%	211.78
K	1476	230.84	1.65%	224.78
L	383	130.59	2.99%	129.77
M	1770	192.78	1.70%	191.90
O	2623	208.13	1.13%	211.57
P	859	164.05	2.47%	163.14
Q	307	197.38	3.83%	191.15
R	714	156.57	2.13%	153.77
전체	2021	181.08	0.74%	180.88

5.4 직종세분류별 평균임금 추정

(1) 산업소분류별 평균임금에 대한 소지역추정법 적용 결과

<표 5-10> 산업소분류별 평균임금에 대한 직접추정량과 소지역추정 기법의 비교

산업대분류	산업중분류	표본	직접추정량	소지역추정
15 음식료품 제조업	151 고기, 과일, 채소 및 유지 가공업	147	118.07	120.45
	152 낙농제품 및 아이스크림 제조업	35	204.67	199.55
	153 곡물가공품, 전분 및 사료 제조업	45	158.10	158.22
	154 기타 식품 제조업	200	148.63	149.29
	155 음료 제조업	83	169.66	167.41
17 섬유제품 제조업 (봉제의보 제외)	171 제사 및 방적업	73	143.93	140.86
	172 직물 직조업	137	129.32	133.23
	173 편조업	22	121.56	123.11
	174 섬유 염색 및 가공업	109	157.84	156.62
	179 기타 섬유제품 제조업	75	135.38	133.65
24 화합물 및 화학제품 제조업	241 기초화합물 제조업	165	223.51	218.98
	242 의약품 제조업	149	202.76	204.67
	243 기타 화학제품 제조업	199	188.40	191.29
	244 화학섬유 제조업	28	231.79	221.68
31 기타 전기기계 및 전기변환장치 제조업	311 전동기, 발전기 및 전기변환장치 제조업	59	162.99	160.39
	312 전기공급 및 전기제어장치 제조업	55	168.61	163.52
	313 절연선 및 케이블 제조업	47	164.82	161.50
	314 축전지 및 일차전지 제조업	24	157.36	153.66
	315 전구 및 조명장치 제조업	39	140.32	144.18
	319 기타 전기장비 제조업	38	145.26	151.62
32 전자부품, 영상, 음향 및 통신장비 제조업	321 반도체 및 기타 전자부품 제조업	494	168.70	168.00
	322 통신기기 및 방송장비 제조업	179	184.50	182.53
	323 방송수신기 및 기타 영상, 음향기기 제조업	139	173.32	174.18
33 의료, 정밀, 광학기기 및 시계 제조업	331 의료용 기기 제조업	45	172.70	170.25
	332 측정, 시험, 항해 및 기타 정밀기기 제조업 (광학기기 제외)	44	180.88	181.55
	333 안경, 사진기 및 기타 광학기기 제조업	34	146.53	138.07
	334 시계 및 시계부품 제조업	17	130.40	148.26

(2) 직종세분별 평균임금에 대한 소지역추정법 적용 결과

<표 5-11> 직종세분류별 평균임금에 대한 직접추정량과 소지역추정 기법의 비교

직종중분류	직종소분류	직종세분류	표본	직접추정량	소지역추정
09 운전 및 운송 관련직	091 선박.항공기 조종 및 관제 관련직	0911 선장, 항해사 및 기관사	30	290.78	290.92
		0912 항공기 조종사	9	474.05	429.96
		0913 관제사	4	223.6	309.27
	092 철도.지하철 기관사 및 관련직	0921 철도 및 지하철 관리자	39	248.52	251.38
		0922 신호원 및 수송원	17	215.47	220.06
	093 자동차 운전 관련직	0931 택시운전원	379	112.94	112.79
		0932 버스 및 승합차 운전원	429	144.88	148.27
		0933 소형트럭 운전원	132	143.77	145.35
		0934 대형트럭 및 특수차 운전	212	174.76	173.54
		0939 기타 자동차 운전원	48	149.93	151.62
	094 크레인 및지게차운전 관련직	0941 크레인/호이스트 운전원	63	189.33	186.91
		0942 지게차 운전원	94	157.76	158.33
	095 운송관련 단순직	0951 선박 갑판원	9	180.40	172.64
		0952 하역관련 노무자	75	138.59	136.66
		0959 기타 배달 및 수하물운반	231	135.23	135.18
22 환경.인쇄 .목재.가구 .공예 및 생산단순직	221 환경공학기술자	2211 환경공학기술자	22	205.11	210.36
		2212 환경 및 보건위생검사원	23	204.50	200.06
	222 품질, 비파괴, 안전 공학관련직	2221 비파괴검사원	9	183.00	190.04
		2222 산업안전 및 위험관리원	81	211.81	211.01
		2229 그외기타공학관련종사자	5	317.64	296.92
	223 환경관련 장치 조직원(상하수, 소각)	2231 상하수처리 관련 조직원	34	205.58	197.39
		2232 소각로 관련 장치 조직원	17	152.53	155.5
	224 인쇄 및 사진현상	2241 인쇄기조직원	81	163.6	165.75
		2242 사진인화 및 현상 조직원	7	186.34	195.64
	225 목재, 펄프, 종이가공 및 제조 관련 조직원	2251 목재가공 관련 조직원	20	127.5	132.95
		2252 펄프/종이제조 및 마무리 관련 조직원	21	167.73	167.78
		2253 종이제품제조 관련조직원	28	162.71	159.57
		2254 목재펠프종이 가공관련 등급원 및 검사원	3	108.86	118.86
	226 가구, 간판제작, 공예원 및 기타 제조 관련직	2261 가구조립 및 검사원	22	121.81	129.69
		2262 가구제조/수리원	52	135.51	138.52
		2263 악기수리원 및 조율원	2	154.54	145.62
2264 공예원		6	87.06	106.48	
2265 귀금속 및 보석 세공원		19	170.80	159.94	
2266 간판제작원		7	170.43	166.4	
2269 기타 제조 관련직		31	123.92	124.25	
227 생산관련 단순직	2271 생산관련 단순노무자	262	82.00	83.15	
	2272 포장원	153	90.11	90.74	

제 6 장 향후 연구방향(2004년 연구계획)

2003년 연구는 산업소분류별 평균임금 산출하는데 소지역 추정법의 적용가능성을 제조업 중에서 일부분에 대해서 일반화 회귀모형을 적합하여 검토하였으며 중앙고용정보원의 자료를 보조정보로 사용할 수 있는지에 대한 가능성을 분석하였으나 분류코드체계의 불일치 등으로 직접 이용하는 것이 쉽지 않다는 점을 파악하였다. 분류코드의 전환체계를 연구한다면 산업소분류별 평균임금을 산출하는데 보조정보로 이용할 수 있을 것으로 판단된다. 따라서 산업분류코드를 상호비교 전환할 수 있는 전환체계는 해당분야 전문가들의 자문을 통해서 개발하는 연구를 추진할 필요가 있다. 산업소분류별 평균임금을 소지역 추정법을 이용하여 산정하기 위해서 2003년에 연구한 일반화회귀모형을 모든 산업소분류에 대해서 적합하고 이를 실제 통계생산에 활용하기 위해서는 SAS 또는 C등의 프로그래밍언어를 이용하여 패키지 형태의 프로그램을 작성하는 연구가 필요할 것이다.

직종세분류별 평균임금산출을 위한 소지역 추정기법에 대해서는 2003년에 심층적인 연구가 진행되지 않았지만 소지역 추정법에서 사용할 모형으로는 산업소분류에서 개발한 일반화회귀모형을 직종세분류별 평균임금산출 소지역 추정법의 방안으로 검토하는 것이 효과적이라 생각된다. 직종세분류별 소지역 추정법을 적용하는데 유용하게 사용될 수 있는 외부보조정보에 대한 탐색연구도 2004년에 추진해야할 주요한 과제일 것이다. 2004년에 연구되어야할 핵심적인 주제들을 아래와 같이 요약할 수 있다.

- (1) 2003년도 임금구조기본통계조사 자료의 심층적 분석
- (2) 소지역 추정법에서 보조정보로 이용할 수 있는 외부통계자료 탐색 및 분석
- (3) 전체 산업소분류별 평균임금과 취업자를 추정하기 위한 산업분류체계의 집락화분석
산업소분류별로 통계를 생산하는데 모집단의 규모 또는 하위범주수의 과소 등으로 유사한 소분류를 몇 개 묶어서 그룹핑하여 소지역추정법을 적용할 수 있도록 집락화하는 연구
- (4) 산업소분류별 평균임금의 추정치를 산출하고 이에 대한 표본오차도 계산하는 방법과 프로그램 개발연구 : 표본오차의 계산방법으로 잭나이프법과 일반화된 분산 함수(GVF : generalized variance function)를 이용하는 방법 연구
- (5) 모든 직종세분류별 평균임금의 산출하는 소지역 추정법의 일반화회귀모형을 결정

(6) 직종세분류별 평균임금과 취업자 수의 추정치와 추정오차를 계산하는 방안연구 및 계산하는 프로그램 개발

(7) 미국 OES에서 벤치마킹하는데 기준으로 이용하는 보조정보와 벤치마킹 승수를 계산하는 과정 및 알고리즘을 연구하고 우리나라에서 적용할 수 있는지 가능성 연구

(8) 표본오차를 계산하는 방법으로 일반화된 분산함수를 이용하는 방안 연구 및 알고리즘 개발

(9) 산업소분류별이나 직종세분류별 평균임금과 취업자 산출을 위해서 채택된 소지역 추정법의 모형과 표본오차를 계산하는 알고리즘의 타당성과 신뢰성을 수치적으로 검증할 수 있는 방법으로 인위적인 모집단을 시뮬레이션으로 생성하여 추정치와 인위적인 모집단의 참값을 비교 분석하는 방안 연구.

참 고 문 헌

- [1] 이계오(2000) 시군구 실업자 추정을 위한 소지역 추정법, 응용통계연구, 제13권 2호, 275-286
- [2] 한국통계학회 (2002). 임금구조통계조사 표본설계 최종보고서.
- [3] Battese,G.E., and Fuller,W.A.(1981) Prediction of county crop areas using survey and satellite data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 500-505
- [4] Battese,G.E., Harter,R.M., and Fuller,W.A.(1988) An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36
- [5] Bureau of the Census (1978) The current population survey: *Design and Methodology*, Technical paper 40, Washington, D.C.: Richard Tiller
- [6] Burtless, G., and Vroman, W.(1984) The performance of unemployment insurance since 1979, *paper presented at the December 1979 meeting of the Industrial Relations Research Association*, Dallas, Texas
- [7] Corson, W., and Nicholson, W.(1988) An examination of declining UI Claims during the 1980s, *Mathematica Policy Research Draft Report*
- [8] Cowles,M.K. and Carlin,B.P.(1996) Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91, 883-904
- [9] Datta, G. S., Day, B. and Basawa, I. (1999) Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference*, 75, 269-279
- [10] Datta, G. S., and Ghosh,M.(1991) Bayesian prediction in linear models: Applications to small area estimation, *The Anals of Statistics*, 19, 1748-1770
- [11] Datta, G. S., Lahiri, and Maiti, T.(1999) Empirical Bayes estimation of median income of four-person by states using time series and cross-sectional data, *Technical Report, Department of Statistics, University of George-Athens*
- [12] Dempster, A. P., and Tomberlin, T. J. (1980) The analysis of census undercount from a post-enumeration survey, *in Proceedings of the Conference on Census Undercount*, pp. 88-94
- [13] Erickson, E. P.(1974) A regression method for estimating populations of

- local areas. *Journal of the American Statistical Association*, 69, 867–875
- [14] Fay, R. E., and Herriot, R.(1979) Estimates of income for small places: An application of James–Stein procedures to census data, *Journal of the American Statistical Association*, 74, 269–277
- [15] Ghosh, M. and Lahiri, P.(1987) Robust empirical Bayes estimation of means from stratified samples, *Journal of the American Statistical Association*, 82, 1153–1162
- [16] Ghosh, M., Natarajan, K., Kim, D. and Waller, L.A. (1997) Hierarchical Bayes GLM's for the analysis of spatial data: an application to disease mapping. *Tech. Rep.* , University of Florida–Gainesville
- [17] Ghosh, M. and Rao, J.N.K (1994) Small area estimation: an appraisal. *Satistical Science*, 9, 55–93
- [18] Gonzalez, M.E. (1973) Use and evaluation of synthetic estimates, *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 33–36
- [19] Henderson, C. R. (1950). Estimation of genetic parameters (Abstract), *Annals of Mathematical Statistics*, 21, 309–310.
- [20] Hobert, J.P. and Casella, G. (1996) The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91, 1461–1479
- [21] Jiang, J. (1996) REML estimation: asymptotic behaviour and related topics. *Annals of Statistics*, 24, 255–286
- [22] Jiang, J. , Lahiri, P.A. and Wan, S.(1998) Jackknifing the mean squared error of empirical best predictor, *Technical Report, Department of Statistics, Case Western Reserve University*
- [23] J. N. K. Rao(1999) Some recent advances in model-based small area estimation, *Survey Methodology*, Vol25, No.2, 175–186
- [24] Lahiri, P.A. and Rao, J.N.K. (1995) Robust estimation of mean squares error of small area estimators. *Journal of the American Statistical Association*, 82, 758–766
- [25] Leslie Kish(1965), *Survey Sampling*, John Wiley & Sons Inc., New York
- [26] MacGibbon, B., and Tomberlin, T. J. (1989) Small area estimates of proportions via empirical Bayes techniques, *Survey Methodology*, 15, 237–252

- [27] Malec, D., Sedransk, J., and Tompkins, L. (1993) Bayesian predictive inference for small areas for binary variables in the National Health Interview Survey, *in Case Studies in Bayesian Statistics*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla, New York:Springer-Verlag, pp. 377-389
- [28] Manas, Lahiri, Michael Larsen, and John Reimnitz(1999) Composite estimation of drug prevalences for sub-state areas, *Survey Methodology*, Vol.25, No.1, 81-86
- [29] Morris H. Hansen, William N. Hurwitz, and William G. Madow(1993), *Sampling Survey Methods and Theory Vol I &II*, John Wiley & Sons Inc., New York
- [30] Moura, F. and Holt, D. (1999) Small area estimation using multilevel models. *Survey Methodology*, 25, 73-80
- [31] Nandram, B., and Sedransk, J.(1993) Bayesian predictive inference for a finite population proportion: Two-Stage Cluster Sampling, *Journal of the Royal Statistical Society, Ser. B*, 55, 399-408
- [32] Prasad, N.G.N. and Rao, J.N.K. (1990) The estimation of mean squared errors of small area estimators. *Journal of the Americal Statistical Association*, 85, 163-171
- [33] Purcell, N.J. and Kish, L. (1979) Estimation for small domains. *Biometrics*, 35, 365-384
- [34] Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New York.
- [35] Shao,J., and Tu,D. (1995) *The Jackknife and the Bootstrap*. New York:Springer-Verlag
- [36] Singh, M.P., Gambino, J. and Mantel, H.J. (1994) Issues and strategies for small area data. *Survey Methodology*, 20, 3-22
- [37] Stroud, T. W. F. (1987) Bayes and empirical Bayes approaches to small area estimation of small area statistics, *in International Symposium on Small Area Statistics*, eds. R. Paltek, J. N. K. Rao, C.E. Saerndal, and M. P. Singh, New York: Wiley, pp. 124-140
- [38] William G. Cochran(1977), *Sampling Techniques 3rd ed.*, John Wiley & Sons Inc., New York
- [39] You, Y. and Rao, J. N. K. (1999) Pseudo Bayes small area estimation using a simple random effects model and sampling weights. *Tech. Rep.*, Statistics Canada.