

**「2011년 통계자료에디팅(SDE) 국제회의」**

# **국외출장 결과보고**

**2011. 6.**

**통계개발원**

# I. 출장 개요

## 가. 출장배경

- UNECE가 주관하고 개최하는 통계자료 에디팅(Statistical Data Editing) 회의에서는 국가통계기관에서 생산되는 통계자료의 에디팅과 관련된 경험교환, 새로운 기법 개발, 방법론적 지침 문제를 다루고 있어 에디팅 분야의 선진 동향을 파악하고 이 분야 전문가와 의견을 교류할 수 있는 좋은 기회가 됨

## 나. 출장목적

- 본 회의에 참석하여 에디팅 업무 및 연구에 대한 최근 동향을 파악하고 에디팅 업무와 에디팅 방법론 연구를 수행함에 있어 질적 향상을 도모하고자 함

## 다. 출장자

- 조사연구실 이의규 사무관, 정보화기획과 이준휘 주무관

## 라. 출장기간

- 2011. 5. 8. ~ 5. 13. (4박 6일간)

## 마. 출장지

- 슬로베니아 류블라나

## 바. 출장일정

구 분		시 간 대	주요내용
1일차	5.8(일)	13:00~17:45 20:20~21:35	- 이동(인천공항→프랑크푸르트) - 이동(프랑크푸르트→류블라나) - 호텔 이동
2일차	5.9(월)	09:00~17:15	- 회의 참가: 행정/센서스 자료 에디팅 등
3일차	5.10(화)	09:00~17:30	- 회의 참가: 매크로/마이크로 에디팅 등
4일차	5.11(수)	09:00~15:30	- 회의 참가: 최신 에디팅 방법 등
5일차~ 6일차	5.12(목)~ 5.13(금)	12:15~13:40 19:45~13:05	- 귀국(류블라나→프랑크푸르트) - 귀국(프랑크푸르트→인천공항)

## II. 회의 개요

### 가. 회의명

- 통계자료에디팅에 관한 UNECE 국제회의

### 나. 회의목적

- 국가통계기관에서 생산되는 통계자료의 에디팅과 관련된 경험교환
- 새로운 기법 개발 및 방법론적 지침 제시

### 다. 회의일시 및 장소

- 2011. 5. 9. ~ 5. 11. (3일간)
- 슬로베니아 류블라나

### 라. 참가자 : 약 80명

- 호주, 오스트리아, 캐나다, 크로아티아, 핀란드, 프랑스, 독일, 헝가리, 이스라엘, 이탈리아, 멕시코, 네덜란드, 뉴질랜드, 노르웨이, 한국, 러시아, 슬로바키아, 슬로베니아, 스페인, 스웨덴, 스위스, 타지키스탄, 영국, 미국 등

### 마. 주요의제

- 행정자료와 센서스 자료의 에디팅
- 전자 수집 자료의 에디팅
- 매크로 에디팅 방법
- 마이크로 에디팅 방법과 소프트웨어
- 조직문화의 변화
- 국제협업
- 최신방법
- K-base/wiki

### Ⅲ. 주제별 발표내용

#### 가. 슬로베니아 환영사

- 슬로베니아의 등록자료 활용 역사는 30년, 최근 센서스는 완전한 등록기반 센서스였으며 약 4천만 유로를 절약하였다고 함
- 행정자료에 많이 의지하게 됨에 따라 에디팅 절차는 특별한 관심이 요구되고 있으며 특이치를 다루는 데 있어 많은 도전에 직면
- 효율성을 위한 일반화된 에디팅 시스템과 소프트웨어에 특별한 관심을 가짐

#### 나. 주요 내용

##### 1) 행정 자료와 센서스 자료의 에디팅 (Editing and administrative and Census data)

- ▶ 2010년 이탈리아 농업총조사를 위한 자료 에디팅과 임퓨테이션의 주요 쟁점(The main innovations of data editing and imputation for the 2010 Italian Agricultural Census):
  - 총 센서스 오차를 줄이기 위해 새로운 E&I 시스템이 이행됨을 소개
  - DIESIS(Data Imputation Editing System-Italian Software)와 CONCERT 에디팅 소프트웨어에 대해 설명
  - 새로운 E&I 전략이 센서스 자료의 품질 향상 도모
- ▶ 2011 영국 센서스: 에디팅과 임퓨테이션 절차의 개관(2011 UK Census: An overview of the edit and imputation process):
  - 영국 센서스 에디팅 전략과 2011 센서스의 개관에 대해 설명
  - E&I를 위해 CANCEIS를 이용하고 있음
- ▶ 2011 슬로베니아 등록 기반 센서스에서의 에디팅 절차(Editing process in the case of Slovenian register-based Census):

- 첫 등록기반 센서스의 자료처리 절차를 소개
- 자료 로딩, 자료통합, 자료수정, 결측치 대체로 구성됨
- ▶ 2010 스위스 인구 센서스의 가구 특성 대체(Imputing household characteristics in the register-based survey of the Swiss population Census):
  - 첫 등록기반 센서스 시행
  - 자치단체별 가구통계를 생산하기 위해 가구 특성을 대체
- ▶ 프랑스 INSEE의 행정자료와 조사자료의 통합 이용을 기초로 한 통계자료의 품질향상(Quality Improvement of individual data and statistical outputs based on combined use of administrative and survey data):
  - 영향치를 검출을 위한 선택적 에디팅 절차를 이에 응용
- ▶ 캐나다 통계청에서의 국세자료의 처리 절차(Processing Methodology of Tax Data at Statistics Canada):
  - 소득세, 사업소득세, 부가가치세 자료의 처리절차의 개관과 당면 과제를 발표
- ▶ 영국의 연간 경제조사에서의 에디팅의 효율성 제고를 위한 부가가치세 자료 이용의 편익 평가(Evaluating the benefits of using VAT data to improve the efficiency of editing in a multivariate annual business survey):
  - 부가가치세를 이용하여 연간경제조사의 주요변수의 예측치 생성
  - 예측치를 이용하여 내검규칙을 개선하고 선택적 에디팅 (스웨덴 프로그램 SELEKT 이용)에 기댓값으로 적용
- ▶ 행정자료와 조사자료의 결합: 경제조사에서 에디팅과 대체의 잠재적 이익과 효과(Combining administrative and survey data: potential benefits and impact on editing and imputation for a

structural business survey):

- 이탈리아 중소기업의 사업체 조사 재설계 프로젝트를 다룸
- 행정자료의 정확성과 이용가능성을 탐색적 분석방법으로 평가

## 2) 전자 수집 자료의 에디팅 (Editing of electronic collections)

▶ 캐나다 인구센서스 인터넷조사에서의 에디팅의 진화(The evolution of edits in the Canadian Census of Population online questionnaires):

- 온라인 자료수집 부문에서 지난 10년간의 교훈을 발표
- 종이조사보다 불응률이 낮음
- 응답률을 높이기 위한 센서스 홍보 전략, “wave 방법” 소개

▶ 매출액 통계에 사용되는 다양한 자료의 에디팅(Editing mixed mode data for turnover statistics):

- 매출액 통계에 부가가치세를 활용
- 부가가치세에 대한 에디팅 패턴 소개와 그에 따른 자동 에디팅

▶ 스페인 통계청의 온라인 자료 수집 시스템 소개(Towards a corporate-wide electronic data collection system at the National Statistical Institute of Spain):

- 에디팅 업무 감소 등 온라인 자료수집의 장점 강조
- GSBPM을 활용한 자료수집 시스템 구축 프로젝트, “INE 프로젝트” 소개
- 통계조사에 사용되는 절차를 재사용하고 다양한 조사 자료를 통합하는 것을 목적으로 함

## 3) 매크로 에디팅 방법 (Macro editing methods)

▶ 호주의 이상치 검출과 우선순위 매김을 위한 선택적 매크로 에디팅

(A macro significance editing framework to detect and prioritize anomalous estimates):

- 이상치를 검출하기 위한 유의성 판단기준에 근거한 점수(score)를 사용하는 매크로 에디팅의 프레임 워크를 설명
- 계층적 매크로 에디팅을 조사 자료에 적용하여 소개
  
- ▶ 네덜란드의 매크로 에디팅을 위한 일반 소프트웨어 패키지 (MacroView: a generic software package for developing macro-editing tools):
  - 사용자 맞춤형 매크로 에디팅 툴 소개
  - 집계, 자료변환, 데이터 시각화, 표와 도식의 재생산 등의 기능을 포함
  - 현재 경제조사에 사용 중
  
- ▶ 뉴질랜드 business performance survey에서의 매크로 에디팅 도입에 의한 효율성 제고(Improving efficiency by introducing macro editing in New Zealand business performance survey):
  - 효율성 제고를 위해 뉴질랜드 통계청은 자동 에디팅 확대 적용과 선택적 에디팅 사용
  - 최근 개발된 에디팅과 임퓨테이션에 대한 quality indicator의 검토 결과를 기술
  
- ▶ 경제통계 조사의 공통 에디팅 전략과 자료 처리(The Common Editing Strategy and the data processing of Business Statistics surveys):
  - 캐나다 통계청은 2010년 품질을 보증하고 응답률을 높이기 위해 다양한 경제 조사를 하나의 프레임 워크로 통합하고 있음
  - 대부분의 경제조사를 위한 공통의 에디팅 전략(CES)의 개발에 대해 설명

- ▶ 캐나다 통합 경제 통계 프로그램에서 top-down 방식을 이용한 품질 및 효율성 제고(Achieving Quality and Efficiency using a top-down approach in the Canadian Integrated Business Statistics Program):
  - top-down 방식은 새로운 반복적 자료 처리과정으로서 추정치 뿐 아니라 quality가 채택 수준에 도달할 때까지 quality indicator가 분석됨
  - 얻어진 quality indicator와 목표값과 비교하여 시스템은 자료 수집이 중단할 수 있는 지 등을 결정
- ▶ 스페인의 최적화 문제로서의 선택적 에디팅(Selective editing as a stochastic optimization problem):
  - 선택적 에디팅의 문제를 이론적 틀에서 다룸
  - 좋은 선택적 에디팅 전략을 최적화 문제로 해결
- ▶ 스웨덴 국제무역 통계의 선택적 에디팅(Selective editing in the international trade in services):
  - 스웨덴 통계청은 국제무역통계의 선택적 에디팅 툴인 SELEKT의 적용을 발표
  - SELEKT는 어떤 오류자료가 통계에 가장 크게 영향을 주는 지를 결정하여 가장 중요한 에러를 우선적으로 재접촉하여 응답부담과 비용을 최소화함
- ▶ 영국 경제조사에 대한 에디팅의 효율성 제고(Improving the efficiency of editing for ONS business surveys):
  - 경제조사에 대한 에디팅 절차의 효율성을 제고하기 위해 Eden 프로젝트에 대해 발표
  - 자료의 품질을 유지하면서 응답부담을 줄이고자 함
  - 선택적 에디팅 SELEKT를 연간 경제조사에 적용한 사례를 보임



#### 4) 마이크로 에디팅 방법과 소프트웨어 (Micro editing - methods and software)

- ▶ 오스트리아의 R 패키지 VIM(Imputation of Complex Data with R-Package VIM: Traditional and New methods based on Robust Estimation):
  - 오스트리아 통계청은 자료에서 무응답자료의 구조를 분석·탐색하고 고품질 그래프를 제공하며 다양한 방법을 통해 무응답을 대체하는 R 패키지 VIM(Visualization and Imputation of Missing values)에 대해 발표
  
- ▶ 이탈리아의 R 패키지 StatMatch와 조사 자료의 통계적 매칭과 임putation(Statistical Matching and Imputation of Survey Data with the Package "StatMatch" for the R Environment):
  - 같은 목표 모집단에 대한 두 자료의 통합을 목적으로 함
  - 통계적 매칭을 위한 비모수적 접근방법, 혼합된 방법에 대해 R 패키지 이용과 함께 설명
  
- ▶ 조사 자료에서 다변량 이상치 검출을 위한 소프트웨어(Software for multivariate outlier detection in survey data):
  - 조사자료에서 이상치를 탐색하는 알고리즘을 소개하고 R 패키지 rrcovNA의 이행을 발표
  - 사례 적용에서 도표화와 이상치 판단 틀을 소개
  
- ▶ 캐나다의 경제조사에서의 priority score를 사용한 추적조사 운용(Collection Follow-Up Operation Using Priority Score for Business Surveys):
  - 캐나다 통계청에서는 추적조사 최적 운용을 결정하기 위해 여러 가지 방법의 비교를 위한 시뮬레이션을 실행
  - 경제적으로 가중된 응답률(EWRR)의 최대화와 관련됨

▶ 스웨덴의 에디팅(The Edit):

- 스웨덴 통계청은 유의 에디팅(Significance editing) 또는 선택적 에디팅(Selective editing)을 위한 도구인 SELEKT를 개발(SAS 프로그램과 SAS macro로 작성) · 소개

▶ 독일의 데이터 에디팅 개념의 재설계(Redesigning a data editing concept):

- 독일은 경제통계에 대한 자료 에디팅에 대한 새로운 접근방식에 관심을 가짐
- 새로운 소프트웨어에 의해 수행될 목표들과 에디팅 개념의 재설계를 위한 기본 조건에 대해 설명

5) 조직 문화 변화 (Changing organizational cultures)

▶ 에디팅 업무 구조 변화의 효과(캐나다)(The impact of a changing business architecture on editing):

- 조직 문화가 변화함에 따라 에디팅을 위한 도구모음(tool kit)의 축소 및 기능 개선
- 재사용 가능한 도구를 관리하고 정제
- E&I 기능을 포함한 두 시스템 BANFF, CANCEIS 소개
- 도구의 문서화, 연구 주도의 솔루션 개발 등 강조

▶ 뉴질랜드 통계청의 에디팅과 무응답대체에 영향을 미치는 방향과 시스템 변화(Direction and system changes impacting on data editing and imputation at Statistics New Zealand):

- 2020년까지의 전략계획 소개: “국가통계를 이용한 정보공유 사회 만들기(An informed society using official statistics)”
- 신규통계 개발에 효율적인 gBPM을 도식화하여 소개
- 행정자료 사용의 증가, 측정된 품질을 보다 유용하게 사용

- ▶ 핀란드 통계청의 통계생산을 위한 E&I 프로젝트 결과(Bringing results of the E&I project to the production of Statistics):
  - 핀란드 통계청의 에디팅과 무응답대체 과정 소개
  - 에디팅 방법의 선택, 자료 오류의 종류에 대해 개괄적으로 설명
  
- ▶ 프랑스의 새로운 사업체 통계시스템에 적용된 데이터 에디팅 전략 (First elements relative to the data editing strategy used for the new system of French structural business statistics):
  - 사업체 통계에 사용되는 신규 시스템의 특징 설명
  - 다양한 장치로 입수 자료 비교, 선택적 에디팅 기법 사용 등
  
- ▶ 동적 시스템으로서의 ESS(European Statistical System) 데이터 에디팅 입문: TRIS (A primer in ESS data editing as a dynamic system: TRIS):
  - 데이터 검증을 3단계로 설명, 데이터 에디팅(결측치, 불일치 자료 검출), 무응답대체, 고급 검증(향후 품질향상을 위한 특이치 검출 등)
  - TRIS(Transport Information System)는 정보 시스템 프레임워크로, 자료 전송을 위한 기능을 구현하고 통합함
  
- ▶ 데이터의 증가로 향상된 가용성(네덜란드)(Availability of data: from scarcity to profusion):
  - 경쟁, 양, 질 3가지의 변화: 독점이 아닌 경쟁, 데이터 양의 기하급수적인 증가, 품질의 정의 변화. 자동화, 실시간, 인터넷 기반을 내세운 GPI(구글 물가 지수) 제공
  - 자료를 시각화하고, 자료와 상호작용할 수 있게 하는 구글 차트 도구(Google Chart tool), 두 도시의 생활 물가를 비교할 수 있는 numbeo.com 소개
  - 위험부담을 감수한 획기적인 개선이 필요

- ▶ 보다 효율적인 에디팅 전략 구현을 위한 변화(영국)(Influencing changes to implement a more effective editing strategy):
  - 에디팅의 변화 구현의 잠재적 장애요소 설명
  - 조직 문화를 변화시키기 위해 “Kotter의 변화를 위한 단계” 설명

6) 에디팅의 산업화에 대한 국제협업 (International collaboration on industrialization of editing)

- ▶ 에디팅 산업화를 위한 네트워크 소개(Introduction and presentation from the Network for industrialization of editing):
  - 에디팅 산업화를 위해 GSBPM(Generic Statistical Business Process Model)과 GSIM(Generic Statistical Information Model)을 참조
  - 에디팅 절차를 설명한 GSDEP(Generic Statistical Data Editing Process) 소개
- ▶ GSBPM과 통계 자료 에디팅(Generic Statistical Business Process Model and statistical data editing):
  - GSBPM의 구조와 특성 소개
  - GSBPM과 타 표준과의 관계, SDMX, DDI 등과 연결되어 있음
  - GSBPM은 통계 생산 프레임워크의 일부이며, 정보 모델, 데이터 형식, 다양한 도구와 방법론으로 보완되어야 함
- ▶ 자료의 저장, 공유, 접근과 GSIM과의 연결을 위한 등록 시스템 개선 전략(A strategy to improve the register system to store, share and access data, and its connections to the generic statistical information model(GSIM)):
  - 비용을 절감하고 응답부담을 줄이기 위해 행정자료와 등록 자료를 활용하는 것은 통계생산 프로세스에서 중요함
  - 스웨덴 통계청의 데이터 웨어하우스(DW) 설계도 도식화

- 단일 출처의 마이크로데이터가 아닌 통합 마이크로데이터 처리 절차 설명: 데이터 연결, 유효성 확인 등

## 7) 최신방법 (New and emerging methods)

- ▶ 보정 대체(Partial(donor) imputation with adjustments):
  - 네덜란드와 노르웨이는 공동으로 자료의 일관성을 보장하기 위해 (어떤 항목에 대체된 값은 기존 항목과의 관계에서 어긋난 자료가 대체될 수 있으므로 모든 내검 규칙을 만족할 수 있도록) 대체자료에 최소의 보정을 수행하기 위한 방법을 소개함
- ▶ 이탈리아의 compositional data에 대한 forward search 알고리즘(A forward search algorithm for compositional data):
  - 구성합계 제약을 갖는 자료의 경우에서의 이상치를 탐색하는 forward search 알고리즘을 소개하고 적용
- ▶ 필수가 아닌 내검규칙에서의 자동 에디팅(Automatic editing with soft edits):
  - 네덜란드 통계청에서는 필수규칙과 비필수 내검규칙을 구분할 수 있는 새로운 자동오류위치포착 알고리즘 branch-and-bound 알고리즘을 소개
  - 네덜란드 통계청의 에디팅 시스템 SLICE에 적용
- ▶ 등록 명부에서 이항변수의 대체(On imputation of binary variables in registers):
  - 스웨덴에서는 등록명부의 이항변수의 대체에서 확률대체와 결정적 대체에 대해 정확성과 관련한 사례연구를 발표
  - 이 사례에서 확률대체가 상대적으로 정확성은 높지만 결정적 대체보다 덜 효율적임
- ▶ 이스라엘 센서스에서 행정 파일로부터 소득 대체 및 에디팅

(Imputation and editing of income from the administrative file in the Census)

- 이스라엘은 1995년 센서스에서 소득 자료의 품질분석에 기초하여 2008년 센서스에서는 직업만 묻고 소득 자료는 통계청에서 이용 가능한 국세청 자료로부터 대체함
- 1995년 센서의 소득자료에 대한 분석과 2008년 센서스의 임putation 절차를 소개
- ▶ 선택적 에디팅을 위한 R 패키지(An R package for selective editing based on a latent class model):
  - 이탈리아는 모형 기반의 선택적 에디팅 방법과 이를 실행할 수 있는 R 패키지 SeleMix의 사용에 대해 발표
  - 패키지는 R에서 다운받아 사용 가능

#### 8) 향후 결과물과 K-base/wiki (Future outputs and K-base/wiki)

- ▶ K-base (knowledge base) 웹은 UNECE에 의해 현재 운영되고 있으며 용어 해설 및 자료 제공
- ▶ 점차 온라인 또는 웹 기반 결과물이 요구됨
- ▶ UNECE는 상호 협력할 수 있는 매개체로 wiki를 개발 중

## IV. 시사점 및 참가소감

### 가. 시사점

- 에디팅 분야는 조사방식의 변화에도 불구하고 국가통계기관의 중요한 업무 절차로 인식
- 많은 국가 통계기관이 응답 부담과 조사비용을 줄이기 위해 행정자

료를 사용하고 있으며 통계적 목적을 위한 행정자료의 사용은 통계 생산 과정의 심층적인 검토가 요구되며 국가통계기관으로서는 도전 이자 기회임

- 행정자료는 모두 또는 부분적으로 통계조사자료를 직접적으로 대체 하거나 조사 자료 처리의 효율성을 개선하는 데 간접적으로 사용될 수 있음
- 국가통계작성기관은 행정 자료의 이점과 약점을 분석하고 추가적인 에디팅 및 대체 문제를 다루기 위한 방법론을 개발하여야 함
- 통합된 E&I 시스템의 개발과 E&I 문제를 해결하기 위해 빠르고 효율적인 알고리즘의 사용을 요구함
- 많은 국가에서 수작업 내용검토를 해야 할 대상을 우선순위를 매기 어 자료를 검토하는 선택적 에디팅(selective editing) 기법을 이용
- 새로운 방법 적용에 대한 장벽에는 실무 담당자의 변화에 대한 거리낌과 책임자의 위험 회피도 포함됨
- 새로운 방법에 대한 국제적인 인증 시스템은 변화 도입을 위해 도움이 될 수 있으며 다양한 방법에 대한 장점과 단점에 대한 서류화 는 새로운 방법을 찾고 검증하는 시간을 줄일 수 있음
- EU 일부기관은 외부 사용자에게 지원을 제공하는데 연루되고 싶지 않아 소프트웨어 공유를 꺼림
- 영국은 선택적 에디팅 소프트웨어로 스웨덴의 SELEKT를 사용하고 자동 에디팅과 임퓨테이션을 위해 캐나다의 CANCEIS를 사용, 경제 조사의 효율성 제고를 위해 호주의 선택적 에디팅을 기본으로 하는 등 각 국의 좋은 소프트웨어나 기법을 적극 활용
- 최신 에디팅 방법연구는 공개 프로그램 R을 이용한 연구 발표가 많 았으며 R 패키지가 공개되어 에디팅 방법 및 프로그램 개선에 효율 적임

#### 나. 참가소감

- 이번 회의에서는 이론적 연구 발표, R 프로그램을 활용한 발표, 각국의 에디팅 관련 사례 등 다양한 방법으로 에디팅의 사례와 경험을 공유하는 회의로서 에디팅에 관한 최신 동향을 습득할 수 있는 좋은 기회가 되었음
- 특히 행정자료/센서스의 에디팅 세션은 행정자료의 활용사례가 많아 짐에 따라 이와 관련한 에디팅의 관심이 고조되어 향후 에디팅의 주요 부문이 될 것으로 보임
- 슬로베니아는 인구가 적고 국토가 작은 소국이나 동유럽 국가 중에는 부국에 속하며 통계에 대한 열정과 EU 국제회의에 적극 참여하는 모습이 인상적임

#### 다. 기타

##### 1) 차기 회의 의제

- 선택적 에디팅
- 에디팅에 대한 국제적 해결방안: 표준 방안, 국제협력
- 다중자료나 혼합방식으로부터의 자료통합에서의 에디팅과 대체
- 메타데이터나 파라데이터를 이용한 에디팅 절차의 효율성 분석
- 통계자료 에디팅을 위한 소프트웨어와 도구
- 최신 방법
- 센서스 자료의 에디팅과 응답대체: 등록기반 센서스 중심으로
- 매크로 에디팅

##### 2) 차기 회의는 2012년 9월 말 노르웨이 통계청에서 개최 예정

##### 3) 본 회의에 대한 자세한 정보는 다음 웹사이트에서 얻을 수 있음

[www.unece.org/stats/documents/2011.05.sde.htm](http://www.unece.org/stats/documents/2011.05.sde.htm)