
Eurostat 조사방법론 컨퍼런스(2013 NTS) 참가 결과
2013 New Techniques and Technologies for Statistics(NTS)

2013. 3.

I 개요

1. 출장 개요

○ 출장목적

- 최근 조사 환경의 악화로 인한 자료수집의 어려움을 해결하기 위해 다양한 출처자료의 활용 및 새로운 표본·조사설계 기법과 조사방법에 대한 논의가 지속적으로 대두됨
- 이에 따라 자료 연계, 자료 표준화, 행정자료 및 메타자료 이용방법, 최신의 표본·조사설계 등 혁신적인 방법과 기법 습득
- 2013년 연구과제인 “다양한 자료를 이용한 다목적 표본설계”와 “다양한 출처 자료 처리 및 통계생산 방안”을 수행함에 있어 선진연구동향을 탐색
- 2차 자료 생산을 위한 자료 연계·통합기법 연구를 위해 과제의 핵심이 되는 자료 연계 및 통합, 메타자료, 행정자료 활용, 자료 표준화 등에 신규통계기법을 활용하여 조사방법론 연구의 질적 향상을 도모
- 국가통계방법론에 관한 국제적 동향을 파악하여 통계개발원의 미래 지향적 연구방향을 설정하고 공식통계 방법론과 관련하여 혁신적인 방법과 기법을 습득함으로써, 선진통계기법 조기도입에 활용
- Eurostat 자료 활용 방법 및 국제기구 통계 운영 방법을 습득하고 조사방법론 컨퍼런스의 진행과정을 참관하여 통계청에서 개최하는 인터넷조사 국제워크숍과 국가통계방법론 심포지엄의 운영에 참고

○ 참가회의

- Eurostat 조사방법론 컨퍼런스(2013 NTTS*)

* New Techniques and Technologies for Statistics

○ 참가자

- 통계개발원장 최연옥, 조사연구실 최재혁 사무관,
국제협력담당관실 최규현 주무관

○ 출장지 및 출장기간

- 벨기에 브뤼셀, 2013. 3. 3.(일) ~ 3. 9.(토)

2. 컨퍼런스 개요

○ 컨퍼런스명

- NITS-Conferences on New Techniques and Technologies for Statistics
(The meeting place for research in Official Statistics)
 - Eurostat에서 주관
 - 일반 및 공식통계의 방법론에 관한 최신 연구동향을 교류하고 통계의
장기적인 요구와 발전방향을 논의하기 위한 컨퍼런스
 - 1992년 최초 개최 이후 2~3년마다 비정기적으로 개최됨

○ 컨퍼런스 일시 및 장소

- 2013. 3. 5 ~ 3. 7. (3일간)
- 벨기에 브뤼셀 Charlemagne building

○ 참가자

- 약 400여명 (각국 통계기관, 학계 등 다양한 분야에서 참가)

II 주요 내용

Opening

New Techniques and Technologies for Statistics

✎ Walter J. Radermacher, Chief Statistician of the EU

- 객관적 정책결정의 중요성이 증대되고 있음
 - 빅데이터 등 새로운 자료원이 나타나고 있음
 - 객관적 정책결정을 위한 이론, 체계, 도구 등은 여전히 발전중임
 - 현재 결과물 이상의 정책결정 결과를 얻기 위한 연구가 필요함
- 주위에 산재에 있는 빅데이터를 획득하는 방법 연구 등 국가통계 생산은 새로운 방향으로 나아가야 함

Keynote1

Small Area Estimation: Methods, Application and New Developments

✎ JNK Rao, Distinguished Research Professor at Carleton University

- 설계단계에서의 고려로 인해 직접추정량의 필요성이 줄어들고 있는 추세임
- 관심변수와 관련된 좋은 보조정보는 모형기반 추정량에 좋은 역할을 할 수 있음
 - 국가통계의 통합, 협조 단계에서는 센서스 자료 및 행정자료와 같은 보조자료의 확대가 필요함
- 모형 선택 및 확인은 매우 중요한 단계이고 가능한 외연적인 평가를 하는 것이 매우 바람직함

- 지역단위 자료는 더 쉽게 이용할 수 있기 때문에 지역단위 모형은 넓은 범위를 가지고 있으나 알려진 표본분산의 가정이 제한적임
 - 계층적 베이지안(HB) 방법은 과워플하면서 복합모형에서 다루기 쉬우나, 모수의 사전선택에 주위를 요함(Rao, 2003)
- 모형기반 추정값은 직접 추정값이고 소지역 추정에서의 오차는 대표본의 오차보다 더 명백함
- 모형기반 추정값의 품질 평가를 위한 적절한 기준은 이용자 의도와 충분한 정확한지 여부임
 - 모형기반 추정값이 직접 추정값보다 더 나은 추정값이라 할지라도 받아들이기 충분히 정확하지 않을 수도 있음

Keynote2

Variance estimation by linearization for indicators of poverty and social exclusion in a person and household survey context

✉ Eric Graf, University of Neuchatel

- 테일러 선형화와 일반화 선형화 방법은 조사방법론 분야에서 많이 이용되고 있음
 - 이 방법들은 대표본에서 효율적임(표본크기 1,000 이상)
 - 소표본에서 선형변수들의 근사이론은 이상값에 의해 강건하지 않는 (non-robust) 값이 되기도 함
- 분산추정을 위한 일반화 선형방법은 복합설계에서 주로 사용되나 계산상에서 사용되는 밀도함수를 주의하여 사용해야 함
 - 본 연구에서 분산의 편의를 축소시킬 수 있는 방법을 제안함(특히 가계수입에 관련)

□ Accreditation for transnational research access to official micro-data in Europe

✉ Paola Tubaro, University of Greenwich

○ DwB(Data without Boundaries) 프로젝트 소개

- 유럽에서 국경을 넘는 자료 접근을 향상시키는 프로젝트임
- 실제로 승인과 관련된 장벽의 원인과 실체를 이해하는 것임
- 국가 간 공통점과 차이점을 발견하고 최선의 실행방법을 찾고자 함
- 잠재적인 해결방안을 제안함

○ 자료 승인에 대한 더 동질적인 유럽 모형은 채택하는데 시간이 요구되기는 하지만 실제로는 가능함(자료보유 쌍방의 동의로 시작)

- 다가올 이해당사자들 간의 DwB 협의는 이해가 필요하고 모든 부문의 참여가 기대됨
 - 개선될 ESC-OS* 또는 여권 연구의 확장을 이해하는 것이 매우 중요함
- * ESC-OS: European Service Centre for Official Statistics

□ Micro-data information system MISSY - Metadata for the scientific community

✉ Jeanette Bohr, GESIS - Leibniz Institute for the Social Science

○ MISSY 소개

- 구조화된 설계에서 국가 마이크로자료를 위한 메타자료를 제공하는 온라인 시스템임
- 사회과학을 위한 국가 마이크로자료의 전문적이고 효율적인 사용을 용이하게 함

- 1973년 이후로 가능한 모든 마이크로 센서스 자료를 위한 메타자료를 자세하게 포함하고 있음

- 유럽 마이크로자료를 통합하기 위해 준비중임(DwB)

- 본 메타자료는 다른 시스템의 호환성이 뛰어나

- XLM 메타자료로 가져오기/보내기 실행이 가능함

- 자료모형, 에디팅 방법 등 모든 소프트웨어 소스를 개방함

- 유럽 마이크로자료의 통합의 비밀보호 부문은 Eurostat에서 관리함

Decentralized and remote access to confidential data in the ESS*

* ESS: European Statistical System

☞ Maurice, Destatis

- ESSnet DARA(Decentralized And Remote Access) 프로젝트 소개

- DARA 프로젝트는 2011년 10월에 시작

- 참가국: 프랑스, 헝가리, 영국, 포르투갈, 독일(EC 70% 기금 지원)

- 자료생산 단계에서 원격접속의 높은 보안을 제공하는 것이 목적임

- ESSnet DARA의 예비실험은 Eurostat의 최종단계의 대안적인 해결방안임

Providing remote access to European micro-data

☞ David Schiller, Institute for Employment Research

- DwB 내 WP4(Improving access to OS micro-data) 프로젝트 소개

- 유럽의 원격접속(RA)에 대한 조사

- 유럽의 원격접속 네트워크 의미

- 제한된 네트워크의 시험운용

- 참가국: 프랑스, 네덜란드, 독일, 노르웨이, 영국

○ OS 마이크로자료의 접속 개선

- 신뢰할 수 있는 마이크로자료 접속의 편의
- 마이크로자료의 직접 접속의 제공
- 자료의 흠어짐에 대한 부담 감소

□ Disclosure risk and data utility in flexible table generators

✍ Natalie Shlomo, University of Manchester

○ flexible table 일반서버에 대한 적용방법인 SDC* 비교

* SDC: Statistical Disclosure Control Method

- 이용자들이 record swapping 단계에서 불확실한 매크로데이터에 대해 불편할 수 있으므로 인터넷상의 공표의 문제점과 리스크를 실제적으로 줄일 수 있는 방법임
- record swapping과 제안된 확률적 변동(stochastic perturbation)은 매크로데이터의 실제값과 다른 값을 공표하는 것임

○ 모든 셀을 변동시키는 변동방법(perturbative method)은 더 많이 보호하고 비밀보호를 제공할 수 있음

- SDC 방법의 오류를 피하기 위해 변동방법은 최종 매크로데이터에서 일반서버 내에 “on the fly”를 제공함
- 확률적 변동방법을 사용하는 것은 사용자가 분석결과를 설명할 수 있게 함

○ 향후 가성성과 일치성을 가질 수 있도록 SDC를 향상시키는 연구를 진행 예정

- 표본설계 기반 또는 모형 기반에서의 접속 공개 리스크 등 연구

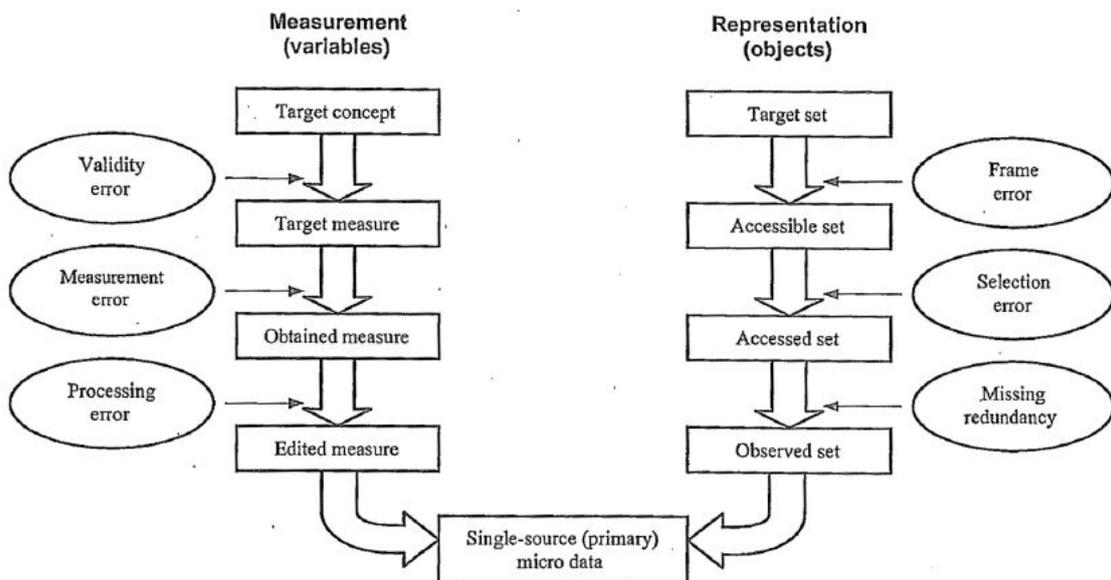
□ Combining survey and administrative data using state space models

✍ Stephen Horn, Australia

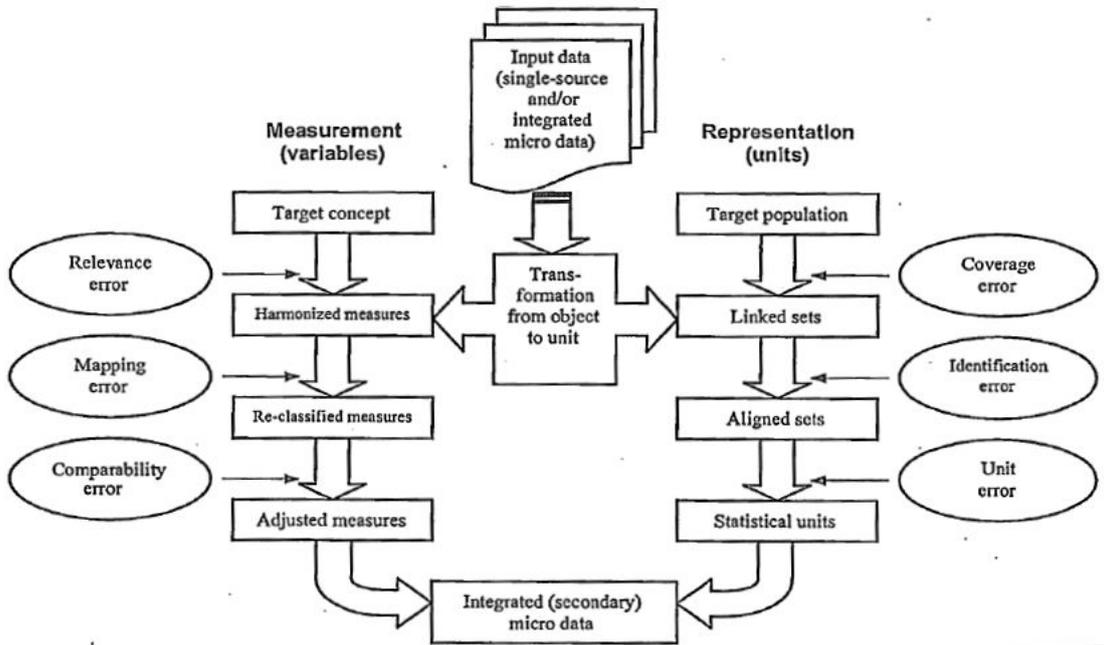
○ 공식통계와 공공정책의 미래

- 밀천은 글로벌한 원인과 개별국의 결과임
- 근원은 오차, 설계, 모형의 도움과 단발업무인지 연속업무인지의 구분임
- 자료 소스를 통합하는 이론으로 주된 임무가 바뀌고 있음
- 빅데이터 알고리즘의 힘과 복잡하고 다양한 포함(coverage)이 강점임
- 보험지불, 산림생태, 복합 반복/패널 조사 등의 경우에 발견
- 새로운 패러다임이 필요한 시점임

○ 1차 자료생산



○ 2차 자료생산



○ 실제의 모니터링 결과

- 연간 모니터링 결과 그래프를 그리기에는 작은 표본크기
- 너무 많은 원격탐사자료
- 적절치 않은 사후층화 추정량
- 빠른 변화에 따른 이동평균의 편이
- 구성 요소가 필요 없는 예측모형
- 모니터링의 빠른 변화를 위한 대안적인 모형기반 추정량이 필요

□ Estimation of measurement error in categorical income survey data

✎ Yury Gubman, Israeli Central Bureau of Statistics

- 수입에 대한 평균 회귀선은 조사자료와 행정자료로 얻어진 자료로 생성

- 분석결과, 불균등 지수는 조사자료가 행정자료보다 더 낮게 나타남
- 지니계수: 조사자료(0.4267), 행정자료(0.4746)
- 회계감사 결과를 토대로 응답자를 분석한 결과, 22%의 응답자가 수입과 그 관련 문항에 대해 정확하게 응답함
 - 이러한 자계식의 정확도가 최종오차를 줄일 수 있음
- 수입과 그 관련 문항에 대한 응답시간은 자영업자들이 고용인보다 더 오래 걸림
 - 자영업자와 고용인들은 수입에 대한 개념적 정의가 달라(실제 측정, 기록된 수입도 그러함) 합동모형으로 분석해서는 안 됨

Session3

New Sampling Methods and Survey Types

□ New sample frames and sample designs for ICT*-survey

* ICT-survey: Information and Communication Technology Survey

✉ Sebastien Faivre, INSEE

○ ICT 조사의 표본추출틀

- 국가 납세자 자료와 지자체 납세자 자료를 통합
- 18세 이상만 대상

○ ICT 조사의 표본설계

- 주거시설(가구)를 추출하여 응답자를 선택하기 때문에 면접조사는 2가지 경우로 구분하여 설계
- 응답자들이 어느 지역에 집중되어 있지 않고 국가 전체에 퍼져있는 조사에서 발생될 수 있으므로 더 이상 ICT 조사만의 문제는 아님
- 계통추출의 한계에서 기대되는 설계효과(design effect)를 높일 수 있는 방법임

- 전국을 두 개의 층(대도시(10,000명 이상), 소도시(10,000명 미만))으로 층화
- 대도시의 경우는 표본추출틀에서 개인을 지역, 성별, 연령을 이용한 계통추출로 추출(단, 한 주거시설에서 한명만 선택되도록 추출)
- 소도시의 경우는 센서스의 조사구를 이용하여 2단 집락추출(단, 집락 내 추출방법은 대도시 추출방법과 동일)

○ ICT 조사의 결과

- 40,000명의 표본에서 전체응답률은 74%임(12%가 주소불일치, 2.3%가 면접조사 불가능/인터넷 사용불가능으로 미조사, 46.3%가 온라인 응답)
- ICT 조사만을 위한 표본추출틀과 표본설계방법의 개발이 필요

□ Data collection with smartphones: experience in a time use survey

✍ Henk Fernee, The Netherlands Institute for Social Research

○ 생활시간조사의 현재 상황

- 조사 난이도가 매우 높고 재조사가 어려움(기억에 의존하는 조사가 아님)
- 낮은 응답률(2006년 조사의 경우 43%)
- 자료 및 결과 이용 빈도가 증가하고 있음

○ 스마트폰 조사 실험 결과

- 생활시간조사 앱은 자료 측정 및 수집에 매우 유용함
- 추가적인 자료를 수집할 수 있는 기회 제공
- 스마트폰 이용자뿐 만 아니라 미이용자도 조사할 수 있음
- 앱을 이용하는 응답자에게 도움이 필요함(음성 또는 가이드라인)

○ 향후 연구

- 스마트폰 조사 경험자와 미경험자의 응답 비교

- 스마트폰 조사 결과와 면접조사 결과 비교
- 면접조사, 온라인조사, 스마트폰조사, GPS조사, 전화조사 등의 자료의 연계·통합 방법

□ Adaptive survey designs using R-indicators*

* R-Indicators: Representativity Indicators

✍ Natalie Shlomo, University of Manchester

- 응답률 또는 포함률은 무응답의 영향을 측정하기에 좋은 측도가 아님
 - 많은 연구에서 자료 수집을 늘리면 응답률이 높아지기 하나 무응답 편의도 증가함을 보여줌
 - 응답과 무응답이 증가함만큼 효과도 증가하는지에 대한 비교도 계속 연구되어옴
- RISQ*: 유럽연합의 7번째 체계연구 프로젝트
 - * RISQ: R-Indicators for Survey Quality
 - 조사에서 모집단 대표성 정도를 측정하는 도구
 - 조사과정 자료와 같은 다른 정보로부터 얻은 인구학적, 사회경제학적 정보를 비교하여 표본단위의 특성을 판단
 - 무응답분석의 질적 판단도구인 R-Indicators를 개발
 - 응답자와 무응답자의 차이의 정도를 측정하는 도구
- R-Indicators와 partial R-Indicators의 이론적 특성은 적응 자료수집임
 - 모의실험 결과, 대표성이 있는 응답자의 부족을 개별적으로 보완하여 응답률이 다소간 상승함을 보여줌
 - 이를 통해 무응답 편의를 다소간 줄일 수 있음
 - 더 대표성이 있는 표본이란 무응답 조정에 의한 최종 조사 가중값의 변동이 적다는 의미이고 이는 더 효율적인 추정량을 제공함
 - 신뢰구간은 넓지만 향후에 정도를 향상시키는 방법을 연구할 예정임

- 횡단자료에서 실제자료로 실험한 결과이며 향후 종단자료도 실험할 예정임

Session4

Small Area Estimation

□ Small area model-based estimators using big data sources

✎ Risto Lehtonen, University of Helsinki

- 소지역 모형에서의 공변량(covariate)을 빅데이터로 활용
 - 빅데이터는 단위(unit)별 자료를 제공하기도 함
 - 결과변수는 소지역모형에서 단위별 자료를 사용하기 위해 연결된 보조변수가 존재해야 함
 - 행정자료, 센서스, 조사자료와 연결된 단위별 빅데이터를 가지는 경우가 드물다는 기술적 도전과 법적 제한에 기인하므로 빅데이터는 주로 소지역 단위별 정보를 이용함
- 빅데이터를 통해 얻은 추정값의 선택편의를 줄이는 데 조사자료를 사용
 - 자료의 부족과 사회적 배제를 직접 측정하기 위해 빅데이터를 사용
 - 빅데이터는 실제로 관심 모집단 전체를 대표할 수 없음(선택 문제)
 - 질적조사에서는 빅데이터와 조사자료 사이의 공통변수의 분포가 다를 수 있음
 - 공통변수가 존재하지 않는다면 분포의 차이를 확인할 수 있는 알려진 상관자료를 이용
 - 빅데이터의 선택편의를 줄이기 위한 가중값 계산 필요
- 소지역 추정값의 타당성을 파악하는데 빅데이터를 이용

- 빅데이터로부터 얻은 자료의 부족과 결핍은 조사자료로부터 얻은 비슷한 정보와 비교
- 빅데이터 추정값과 조사자료 추정값이 일치한다면 부족과 결핍에 대한 증빙을 두 번 확인 가능

□ Small area estimation for victimization data: case study on the violence against women

✍ Michele D'Alo, Istat

- 모든 소지역 추정 방법은 추정값의 신뢰구간을 좁히는 방법임
 - 그러나, MSE 비교에서 볼 수 있듯이 단위별 모형 추정량이 소지역별 모형 추정량이 더 나은 결과를 가짐(직접추정량의 경우보다 약 48%정도 더 우수함)
 - pseudo-EBLUP* 추정량(pseudo-BLUP 추정량보다 더 우수함)이 좋은 결과를 가짐(직접추정량의 경우보다 약 43% 더 우수함)
 - * pseudo-EBLUP: pseudo Empirical Best Linear Unbiased Estimator
 - 국가단위에서는 추정값이 자기 벤치마크(self-benchmark)가 가능하기 때문에 편의를 줄이는 것이 매우 중요하며,
- 로지스틱모형을 이용한 예비결과는 직접추정량이 향상되지 않고 과소추정에 영향을 주고 있음

□ Precision comparison between split and some classical small area estimators

✍ Toky Randrianasolo, The France Institute of Science and Technology for Transport

- 가중값 방법과 벤치마크 보정한 지역추정값 방법에서 착안
- 직접 추정방법
 - 지역단위에서 직접 추정방법은 지역 외 어떤 정보도 사용하지 않는 추정방법임

- 표본단위는 속해 있는 지역의 추정에만 사용됨

○ 제안된 추정량은 가중방법으로 어떠한 관심변수에도 적용 가능함

- 모형의 상수항은 관심변수에 의존함

- 분산계산 관점에서 보완적인 방법임

□ **Treating structural level and variance breaks in a state-space setting: the case of the Dutch road transportation survey**

✉ Oksana Bollineni-Balabay, Maastricht University

○ 9개 지역에 대한 소지역 모형 제안

- 10개의 일변량 모형: 다변량 형태에 유사한 점 추정값으로 잠재적으로 향상된 분산추정값을 생산

- 9차원 모형: 9개 개별 추정값을 이용하여 전체의 추정값을 획득하는 방법으로 선형결합 형태로 추정됨

- 10차원 모형: 9개의 소지역과 국가전체 계열을 이용하는 방법으로 교란항의 공분산행렬의 구조에 대한 적절한 제약을 가지고 있으며, 지역의 통합 오차항 간의 조건부 상관에 대해 공분산이 0이고 추정 정도에 미치는 영향이 거의 없다는 가정을 가지고 있음

○ 모의실험 결과

- 단위와 분산의 불일치와 소규모 표본크기에 대한 문제 해결

- 일변량 모형보다 다변량 모형이 분산이 줄어드는 효과가 있음

- 본 연구결과에서는 10차원 모형이 가장 우수한 것으로 나타남

Session5

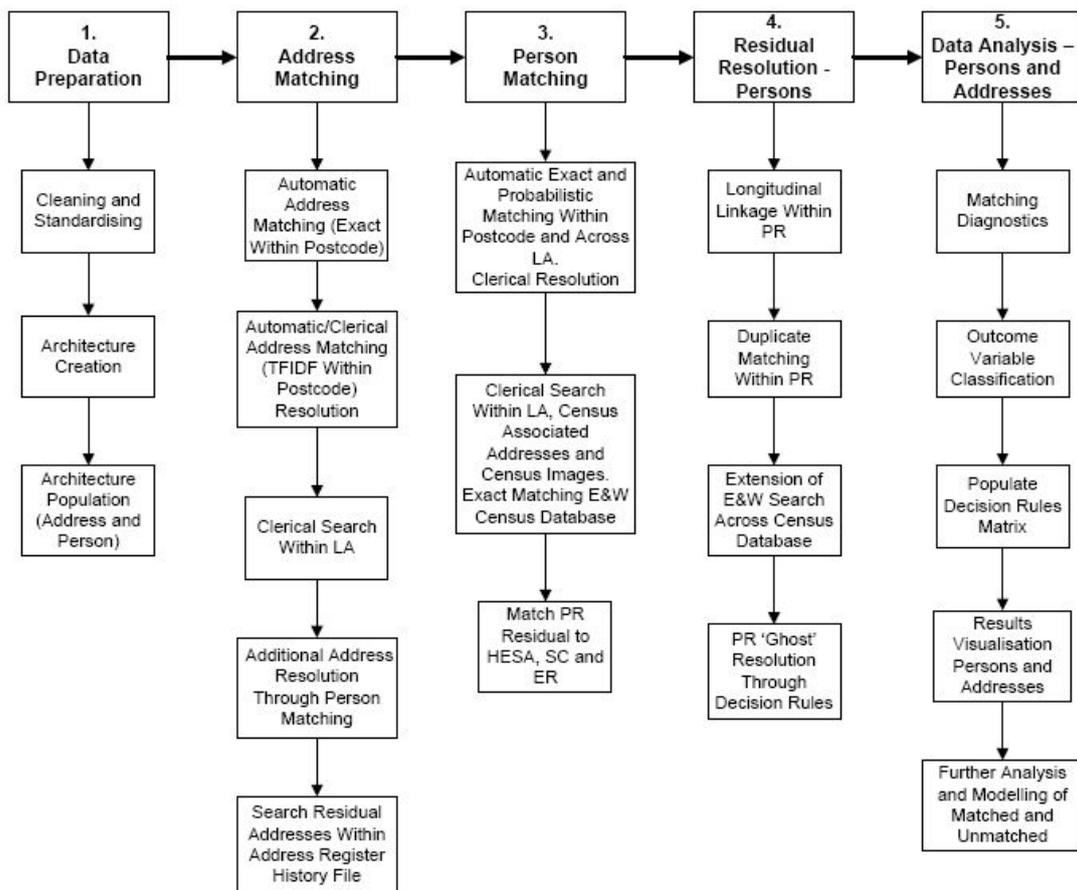
Use of Registers and Administrative Data in Statistics

□ **Matching of census and administrative data for census data quality assurance in the 2011 census of England and Wales**

○ 센서스와 행정자료 연계의 도전

- 자료 관리의 안전과 불확실한 분석에 대한 추가 요구
- 개인자료와 주소자료의 연결과 의미의 복잡성
- 자료의 낮은 활용과 일정하지 않은 품질
- 시의적인 결과에 대한 요구
- 자료관리 상의 연계 작업의 척도 유지
- 연계된 자료의 품질과 일치성의 손상

○ 자료 연계의 과정



○ 추가적인 진행과정 및 연구

- 방법과 체계의 재설정

- 새로운 연계 프로젝트 시행: 센서스 전체 자료 사용

□ A system to estimate levels and growth rates of quarterly turnover

✎ Arnout van Delden, Statistics Netherlands

○ 자료연계 방법론

- 모집단 단위 유형: 기본단위(지역 단위)와 합성단위(기업 단위, 조세 단위 등), 각 자료의 단위들 간의 관계(추정값을 생산하는데 필요)
- 모수 추정: 최초 자료 단위의 결측값 대체
- 방법론: 각 단위의 무응답에 대한 예측이 필요

○ 새로운 시스템의 주요 쟁점

- 자료 재사용에 대한 국가차원의 검토
- MS(major state) 결과의 형태 및 품질의 모니터링 도구
- 성장률 추정값과 단위에 대한 행정자료 사용 방법론

□ Estimating the validity of administrative and survey variables by mean of structural equation models

✎ Sander Scholtus, Statistics Netherlands

○ 추정의 타당성

- 행정자료와 조사자료 간의 비교(단, 조사자료에 오차가 없음을 가정)
- 행정자료의 타당성은 조사자료와 유사함을 가정
- 적용하기 매우 간단하지만 실제로는 조사자료에 오차가 존재함
- 대안적인 방법으로 구정방정식 모형을 제안

○ 모의실험 결과

- 정확모형은 일반적인 방법으로 표본크기와 추정값의 타당성을 판단하는 도구에 의존하고 부분 불규격 측정모형은 지역적 영향만 존재함
- 타당성을 판단하는 다른 도구 대신 실제 타당도 점수 사용
- 범주형 변수: 구조방정식 모형 대신 잠재적 분류모형 사용

□ Dwelling price ranking vs. socio-economic ranking: possibility of imputation

✎ Larisa Fleishment, Israeli Central Bureau of Statistics

○ 연구목적

- 주택가격이 다양한 지리적 단위의 사회경제적 수준에 따라 달라지는지를 탐색
- 지리적 지역 하에서 주택가격의 수준이 계산된 사회경제적 지표를 제외한 거주자의 사회인구학적 특성의 영향이 확장되는지를 탐색

○ 연구결과

- 주택가격 순위는 다양한 지리적 단위의 사회경제적 수준에 영향을 받음
- 본 연구는 지리적 단위에 대한 사회경제적 집단의 수준이 완벽하게 결측일 때(지표로 계산되지 않음) 사용할 수 있는 방법을 제공함
- 이러한 대체방법은 약속된 지표인 사회경제적 지표의 사용자들에게 매우 중요한 방법을 제공함

Session6

Bid Data, Data Mining and Prediction

□ Placing big data in official statistics: a big challenge?

✎ Monica Scannapieco, Istat

○ 연구배경

- 빅데이터: 보통의 소프트웨어와 도구로 처리할 수 없는 자료
- 빅데이터의 처리는 압도적으로 방대한 정보의 경영 시나리오에서 의무적인 책무
- 과정과 체계에서 높은 수준의 새로운 기술과 분석과 통계생산의 새로운 유형이 필요함

○ 통계적 생산과정에서의 빅데이터

- 실제의 조사자료 분석에서의 최근 방법은 높은 품질의 작은 자료의 통합으로 고안한 일반적인 최초방법으로부터 크게 발전되어짐
- 빅데이터가 좋은 기회로 나타나는 동안 빅데이터의 특수성을 해결할 준비가 되어있지 않아 전통적인 방법을 사용해옴
- 전통적인 방법으로 빅데이터 처리가 가능할 수 있는 두 가지 방법은 빅데이터 기술을 도입하는 것과 빅데이터에서 통계적인 분석을 할 수 있도록 방법을 발전시키는 것임

□ Two-stage mining for big statistical micro data

✉ Raja Haddad, SYROKKO, Rossiy, France

○ 통계적인 마이크로자료로 구성된 빅데이터를 관리 및 분석하는 2단계 데이터마이닝 과정

- 첫 번째 단계는 마이크로 자료를 효율적으로 통합하는 것임
- 두 번째 단계는 통합된 자료를 분석하고 시각화하는 단계임

○ 효율적인 통합을 위한 요구사항

- 자료의 개념을 잘 파악하여 적절한 통합방법을 개발하고 결정
- 사용하기 편리한 개념적 특성을 사용
- 새로운 데이터 셋을 만드는 2단계를 효율적으로 구성

○ 자료를 통합하고 Symbolic 시스템 구성을 통한 보급이 기본임

- 이러한 통합은 새로운 분석기법이 필요한 것이 아니라 전통적인 분석

기법을 통해 해석 가능함

- 그러나, 빈도 또는 양적 자료를 이용한 복합통합 하에서는 분석이 어려울 수 있음
- Symbolic 시스템 방법은 이러한 유형의 분석기법에 추천할만 함

□ Techniques for customer behaviour prediction: a case study for credit risk assessment

✎ Laura Maria Badea Stroie, Bucharest Academy of Economic Studies

○ 연구목적

- 경제적 어려움: 새로운 리스크 평가방법을 통한 에이전트의 물리력
- 이론적 배경을 찾기 위한 모형화 방법: 최근의 경제구조 내의 통합을 위한 학술적인 방법을 지양
- 신경망 이론을 이용한 해결방법 모색

○ 신경망 이론의 장·단점

- 일반적인 추정량으로 이상값에 큰 영향을 받지 않으면서 관련 법칙을 찾기 쉬워 빅데이터에 적용하기 적절함
- 가중값을 통해 분포의 종류에 대한 정보를 가지고 있고 일반화할 수 있는 수용력이 높음
- 모형적합에 시간이 많이 사용되며 작동원리를 이해하기 어려움
- “over-fitting” 현상에 노출될 가능성이 높음
- 표준 역증식(standard back-propagation) 알고리즘은 국소최소값 문제에 노출되어 있음

□ Exploiting historical registers: automatic methods for coding c29th and c20th cause of death description to standard classifications

✍ Jamie Carson, University of St Andrews

○ 연구배경

- 19세기와 20세기에 디지털 사망확인서의 사용이 증가함
- 건강, 의료, 유전학 연구에 잠재적으로 의미가 있음
- 150년 동안 사망원인의 변화는 어떠한지 또는 주요 항목에 대한 수기입이 아닌 지동입력방식의 과정은 어떠한지에 대한 방법론적 의미

○ 자료 연계 및 통합 방법

- 일반적인 베이지안 방법, 상호보완적인 베이지안 방법 확률적 하강 경사도 최대가능도(stochastic gradient descent ML) 알고리즘 등 사용
- 오류점검 사전과정을 사용하는 경우와 사용하지 않는 경우로 분류
- 개별적 분류를 통한 통합 방법(최다 빈도, 가장 신뢰성 높은 분류방법, 신뢰한계 등)

○ 연구결과

- 신뢰성 있는 통합방법은 정확도와 일치도 면에서 가장 우수한 전체 자료를 제공
- 자료마다 다른 방법을 적용하는 것이 더 효율적임

Session7

Common Tools

A system for managing process and product quality

✍ Dennis Trewin, Australian Statistician

○ 연구배경

- 다양한 통계생산(조사, 행정, 복합자료 등)에서 이해당사자들의 품질에 대한 주제별 양적 측정이 필요함
- 본 연구에서는 통계결과에 대해 ESS 품질평가 방법을 사용

○ 평가 대상

- 노동력조사, 소비비용지수, 수출입통계, 지방자치단체의 연간계정, 경영 구조통계, 경영 등록자료, 국민 등록자료, 삶의 지표, 국민계정(분기별, 연간 GDP)

○ 평가 항목

- 관련성, 시의성 및 정시성, 비교가능성, 일관성, 접근가능성 및 비교성

○ 평가 결과

- 지난해에 비해 향상되었으나, 측정오차에 대해 높은 리스크를 가지거나 낮은 점수를 가진 항목이 존재함
- 가구조사에서 무응답 편위의 가장 중요한 요인을 찾아야 함
- 평가 연구는 MSE 항목에 더 집중할 필요가 있고 구조 간 더 나은 조직형태를 만들 필요가 있음
- 경제통계에서는 통합과 조직화가 더 필요함
- 통계의 품질에 대한 기록은 과거보다 더 나은 결과를 가져올 것임

□ Integrated statistical systems and their flexibility - How to find the balance?

✎ Rudi Seljak, Statistical Office of the Republic of Slovenia

○ 연구배경

- 빠르고 완전하며, 질적으로 우수하면서 값싼 통계결과에 대한 요구가 증가하고 있음
- 통계생산 결과의 효율성의 최적화 방법에 대한 고민
- IT 도구를 활용한 효율적인 자료수집방법과 관리 및 적용과정이 필요

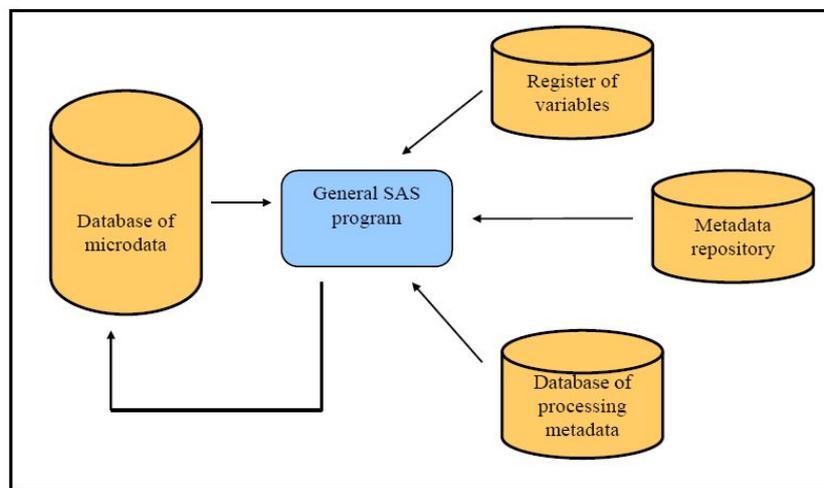
○ 연구결과

- 전체 통합시스템에서 스토브 시스템으로 직접 연결하는 방법은 가장

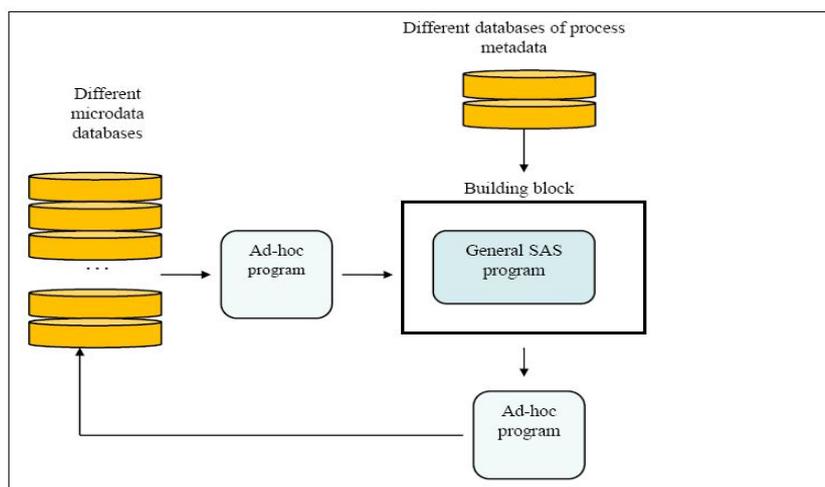
좋은 방법은 아님

- 더 효율적이면서 과정을 체인처럼 쉽게 연결할 수 있는 소규모의 더 유연한 해결방법의 개발이 필요함
- 본 시스템은 발전단계로 일치성에 대한 실험을 실시해야 함

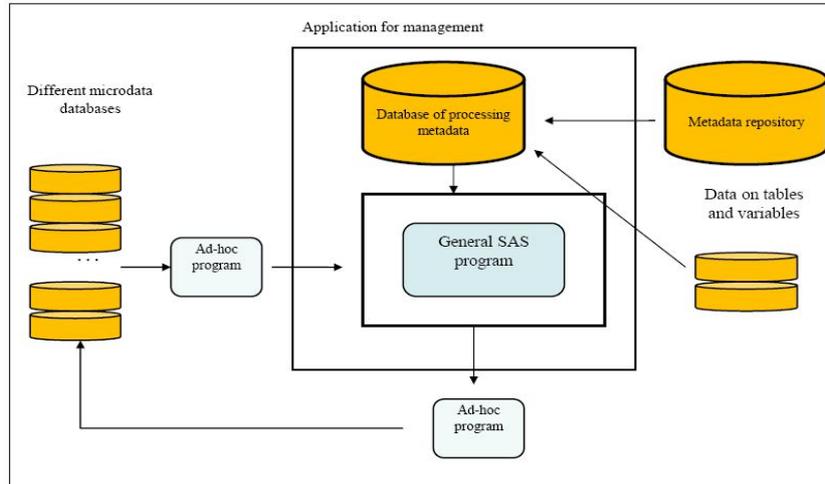
○ 데이터 통합 시스템 과정



○ 데이터 통합 및 분해 시스템 과정



○ 데이터 재결합 시스템 과정



□ Innovation at Statistics Netherlands

✍ Barteld Braaksma, Statistics Netherlands

○ 네덜란드 통계청의 혁신 프로그램

- 결과에 대한 외부의 압박(새로운 결과, 공동연구 등)
- 기술적 도전(빅데이터, 소셜매체 등)
- 내부의 아이디어(권한 부여, 시너지 효과 등)

○ 혁신 프로그램 내용

- 혁신 업무의 로드맵(거친 아이디어→강화된 아이디어→스폰서 탐색→연구제안→개념 증명→실행)
- 이론적인 방법론으로 시작하지 않고 적용단계의 아이디어로 시작
- 좋은 아이디어를 가지는 가장 좋은 방법은 많은 아이디어를 가지는 것을 모태로 함

○ 혁신 시설

- 2012년 5월 개관
- 정교한 아이디어, 브레인스토밍, 워그숍, 열린 IT환경, 의사소통이 용이한 시설

○ 혁신 프로그램의 결과물

- 인터넷에서 얻은 자료 소스, 스마트폰 활용, 빅데이터 활용 등

- 2012년 50개의 아이디어 채택
- 다양한 비통계적 이슈 채택(법적문제, IT 활용 등)

Ⅲ 시사점

□ 유럽연합의 통계연구 및 생산방식의 변화

- 지속적으로 감소하는 응답률과 조사 예산 삭감 등의 조사환경의 악화 속에서 행정자료 활용에 대한 노력 표현
- 최신 정보통신기술을 활용한 자료 수집에 대한 연구 증가
 - 표본조사와 같은 전통조사 방식에서 벗어난 디지털 자료수집에 대한 연구가 활발함
 - 유럽연합의 경우 행정통계 또는 빅데이터 등을 활용한 가공통계를 지향하는 분위기임
- 스마트폰 및 GPS, 네트워크 시스템 등을 활용한 자료수집에 대한 연구가 활발하게 진행
 - 새로운 자료 소스의 활용도는 높으나 모집단 대표성 및 자료 수집의 일관성 부족에 대한 지적이 많고 그에 대한 연구는 초기단계임
- 빅데이터 등 행정자료와 등록자료에 대한 꾸준한 연구가 진행됨
 - 빅데이터, 행정자료로부터 얻은 개인적 성향 및 특화된 계층의 특성을 활용하여 소지역 추정 등에 보조변수로의 활용 가능성이 높음
 - 기존의 정책활용 뿐만 아니라 이용자도 만족시킬 수 있는 이용자 중심의 통계 생산에 활용할 수 있는 방법들은 고민하고 있음

□ 컨퍼런스 준비 및 진행상황 벤치마킹

○ 발표진행 및 자료의 효율적인 운영·관리

- 지난 컨퍼런스 내용을 포함한 인터넷 홈페이지 관리는 효율적으로 되어 있고 발표내용 및 발표자료의 업데이트가 신속하게 처리됨

○ 활발한 포스터 세션 활용

- 점심시간을 활용한 포스터 세션의 경우, 해당 의제를 일목요연하게 정리하여 다양한 내용을 쉽게 접할 수 있는 기회가 됨

○ 발표논문집은 디지털 파일(인터넷과 USB)로 대체

- 회의 일정에 대한 간단한 팸플릿만 제공됐을 뿐 발표문이나 논문 등을 축약한 책자가 없어 발표 내용을 전체적으로 파악하는 데 다소 어려움

○ 기타사항

- 대부분의 발표자들이 발표시간을 지키지 않아, 세션 간 회의실 이동이 원활하지 못했으며 발표 후 질의를 통한 충분한 토론이 진행되지 않음

NTTS 2013 programme | CROS-portal - Windows Internet Explorer

http://www.cros-portal.eu/content/ntts-2013-programme

COLLABORATION IN RESEARCH AND METHODOLOGY FOR OFFICIAL STATISTICS

euostat Login Help Search Advanced Search

Topics ESSnet Research Events News

Home » NTTS 2013 programme

User login

Username

Password

Create New Account
Request New Password

LOG IN

NTTS 2013

- SCIENTIFIC CONTENT
 - NTTS 2013 PROGRAMME
 - NTTS 2013 PRINTABLE PROGRAMME
 - NTTS 2013 PROGRAMME - PRINTABLE OVERVIEW ON ONE A4
 - SESSION 1A - OPENING SESSION
 - SESSION 2B - SPECIAL SESSION ON STANDARDISATION
 - SESSION 2P - POSTER SESSION ON

NTTS 2013 programme

For all papers where full papers have been supplied, the full papers are available via the below links.
Most of the presentations are now online, just click on

	Tuesday 5 March 2013
	Welcome coffee 09:00-09:30 in the lobby at level 2
	Session 1A - Opening session 09:30-12:00 in the "Alcide de Gasperi" room (GASP)
Session chair:	Walter Radermacher (Eurostat)
	Opening address by Walter Radermacher, Director-General, Eurostat
	Research related to Official Statistics in Horizon 2020 , Keynote address by Robert-Jan Smits, Director-General, DG Research and Innovation, European Commission
	Official Statistics and Big Data , Keynote address by Robert M. Groves, Provost, Georgetown University
	Reboot 2020 , Keynote address by Sander Duivestijn, VINT (The Institute for the Analysis of New Technology)
	Session 2R - Special session on Standardisation

※ 자료제공: <http://www.cros-portal.eu/content/ntts-2013-programme>

<부록> 세부회의 주제

Day1 Session 1A : Opening Session
Session 2B : Special Session on Standardization
Session 2P : Poster Session in Indicators, Statistical Disclosure Control and Electronic Data Capture
Session 2Q : Poster Session on Wellbeing, Data Linking and Data Warehouses
Session 3A : Keynote Address
Session 4A : Beyond GDP
Session 4B : Electronic Data Capture
Session 4C : Improving Researchers' Access to Official Micro-data
Session 5A : Data Linking and Data Warehouses
Session 5B : Challenges in Labour Market Statistics
Session 5C : Mobility Statistics

Day2 Session 6A : Keynote Address
Session 7A : Visualization
Session 7B : New Sampling Methods and Survey Types
Session 7C : Special Session on ESSnet Projects
Session 8B : EMOS¹⁾ – Objectives and Next Steps
Session 8P : Poster Session on Spatial and Mobility Statistics
Session 8Q : Poster Session on Small Area Estimation, Respondent Behaviour and Use of Registers/Administrative Data in Statistics
Session 9A : Keynote Address
Session 10A : Small Area Estimation
Session 10B : Response Propensity and Response Burden
Session 10C : Metadata, Data Access and Statistical Disclosure Control
Session 11A : Use of Registers and Administrative Data in Statistics
Session 11B : Geographical Information Systems and Spatial Data Representation
Session 11C : GDP, Prices and National Accounts

Day3 Session 14A : Keynote Address
Session 15A : Big Data, Data Mining and Prediction
Session 15B : Editing, Imputation and Revision
Session 15C : Variance Estimation and Outlier Treatment
Session 16P : Poster Session on Standardization, Editing, Imputation and New Sampling Methods/Survey Types
Session 16Q : Poster Session on Common Tools and Time Series Analysis
Session 17A : Common Tools
Session 17B : Special Session on BLUE-ETS
Session 17C : Special Session on COST²⁾/SHANTI³⁾
Session 18A : Closing Session

1) European Mater in Official Statistics

2) European Cooperation in Science and Technology

3) Sciences, Humanities and Arts Network of Technological Initiatives