

2014 국제 조사방법론 심포지엄
(International Methodology Symposium)
참가 결과보고



2014. 12. 04.

I 개요

1. 출장 개요

○ 출장 목적

- 본 심포지엄은 조사방법론 전반에 관한 워크숍이며, 특히 이번 심포지엄에서는 빅데이터, 자료연계방법, 행정자료 활용, 웹조사 및 패널 조사 등 조사방법론과 관련한 주요 분야에 대해 통계전문가들이 최근 선진 방법론을 토의하는 자리로 다양한 방법론 연구를 위한 좋은 기회가 될 것으로 기대
- 또한 2015년 통계개발원에서 개최하는 국가통계방법론 심포지엄과 관련하여 주요 이슈 및 회의 전반에 관해 벤치마킹

○ 참가회의

- 국제 조사방법론 심포지엄(International Methodology Symposium)

○ 출장자

- 조사연구실 최재혁 사무관, 김경미 주무관

○ 출장기간 및 장소

- 출장기간 : 2014. 10. 28.(화) ~ 11. 2.(일), 4박 6일
- 출장장소 : 캐나다 오타와(Palais des congres de Gatineau)

2. 회의 개요

2014 International Methodology Symposium

Beyond traditional survey taking: adapting to a changing world

- 2014. 10. 29. ~ 11. 1., Canada Ottawa



○ 주요 토픽

- Big data
- Administrative data
- Optimization of data collection
- Electronic questionnaire
- Balanced sampling
- Measurement error
- Record linkage
- Web (panel) surveys
- Mode effects
- Model-based approach
- Microsimulations
- Total survey error

○ 프로그램 구성

Day 1 - 2014.10.29(수)

08:45 - 09:00	Opening Remarks	
09:00-10:00 Plenary Session	Session 1 : Keynote Address Survey Sampling in Official Statistics - Some Thoughts in Directions	
10:00-10:30	Poster Session and Morning Break	
10:30-12:00 Concurrent Sessions	Session 2A : Web Surveys 1	Session 2B : Data Collection 1
12:00-13:30	Lunch	
13:30 -15:00 Concurrent Session	Session 3A : Data Collection Mode Effects	Session 3B : Nonsampling Errors
15:00-15:30	Poster Session and Afternoon Break	
15:30-17:00 Concurrent Sessions	Session 4A : Record Linkage 1	Session 4B : Multiple Sources of Data

Day 2 - 2014.10.30(목)

09:00-10:00 Plenary Session	Session 5 : Waksberg Award Winner Address From Multiple Modes for Surveys to Multiple Data Sources for Estimates	
10:00-10:30	Morning Break	
10:30-12:00 Concurrent Sessions	Session 6A : Total Survey Error	Session 6B : Record Linkage 2
12:00-13:30	Lunch	
13:30 -15:00 Concurrent Session	Session 7A : Big Data	Session 7B : Data Collection 2
15:00-15:30	Afternoon Break	
15:30-17:00 Concurrent Sessions	Session 8A : Microsimulations 1	Session 8B : Web Surveys 2

Day 3 - 2014.10.31(금)

08:30-10:00 Concurrent Sessions	Session 9A : Quality Indicators for Administrative Data	Session 9B : Sampling and Estimation
10:00-10:30	Morning Break	
10:30-12:00 Concurrent Sessions	Session 10A : Alternative Sampling Methods	Session 10B : Microsimulations 2
12:00-13:30	Lunch	
13:30 -15:00 Concurrent Session	Session 11 : Approaches to Inference in Survey Sampling	
15:00-15:30	Session 12 : Closing Remarks	

II 주요내용

1. 캐나다 조사방법론 심포지엄의 주제, 구성 및 운영

- 본 심포지엄은 캐나다 통계청 주관으로 1984년 첫 개최 이후, 2007년 2012년을 제외하고 매년 개최되어 2014년은 29회 개최임
- * 2007년은 워크숍으로 대체, 2012년은 사업체통계 국제 컨퍼런스(International Conference on Business Establishments, ICES) 개최
- 각 국의 통계청 실무자 및 대학교 교수진이 모여 조사 방법론에 관한 세계 연구동향, 최신 통계기법 등을 발표하고 연구 결과 교류

(2014년 이전 최근 5년간 주제)



2013:

Producing reliable estimates from imperfect frames

✍ Keynote Address

Methodological Issues Associated with the Use of Imperfect Frames (Olivier Sautory, Insee, France)

✍ Session

- Multiple Frames
- Two-Phase Sampling
- Alternate Designs
- Using Administrative Data for Frame Creation and Estimation
- Indirect/Adaptive Sampling
- Data Collection Approaches
- Modeling and Estimation
- Alternatives to Censuses
- Web Surveys

- Challenges with Imperfect Frames
- Surveying Difficult-to-Reach Populations
- Frame Development for Censuses and Large Surveys
- Telephone Surveys • Small Area Estimation
- Use of Administrative Data for Business and Agriculture Surveys
- Calibration • Improving Health Surveys



2011: Strategies for Standardization of Methods and Tools - How to get there

✍ Keynote Address

Standardising methods and tools: a look back over 30 plus years of experience (Susan Linacre, ABS, Australia)

✍ Session

- Corporate Business Architecture: Household Survey Frames
- Sampling and estimation • Selective editing
- Harmonization of methods as part of large-scale standardization projects for business surveys
- Dissemination and data access • Confidentiality
- Selective editing methods and tools-an Australian Bureau of Statistics
- Standards and guidelines for the design and testing of Internet questionnaires
- Outliers and imputation • Content and collection
- Confidentiality methods and tools for accessing data while preserving confidentiality
- Using standardized methods and tools for post-collection processing
- Standardized software • Modeling and estimation
- Calendar effects and temporal coherence
- Building and using generalized systems • Framework



2010: Social Statistics: The Interplay among Censuses, Surveys and Administrative data

✍ Keynote Address

Statistics without survey? About the past, present and future of data collection in The Netherlands (Jelke Bethlehem, Statistics Netherlands)

✍ Session

- Small Area Estimation
- Multi-mode Data Collection
- Microsimulations
- Sampling Frames
- Justice Data
- Administrative Censuses
- Multiple Frames
- Poverty Data
- Linkage of Health Data
- Complementary Surveys
- New developments in Record Linkage Methods
- Validation
- Treatment of item nonresponse in social surveys
- Databases
- Use of Auxiliary Sources in Weighting
- Employment Data
- The Canadian Census
- Linkage Methods
- Reconciliation
- Realtime Access to Data from Multiple
- Non-response



2009: Longitudinal Surveys: from Design to Analysis

✍ Keynote Address

The Methodological Research Agenda for Longitudinal Surveys (Peter Lynn, University of Essex, UK)

✍ Session

- Event History Data Collection
- Data Collection and Linkage
- Attrition Bias and Nonresponse Weighting
- Measurement Errors
- Analysis of Longitudinal Survey Data
- Edit and Imputations

- Issues in Economic Surveys · Weighting and Estimation
 - Synthetic Data Approaches to Confidentiality
 - Longitudinal Health Data: Issues and Challenges
 - Data Collections Issues in Longitudinal Survey
 - Accommodating Missing Data in Longitudinal Survey Data Analysis
 - Factors and Impacts of Non-Response
 - Redesign of Large-scale Longitudinal Surveys
 - Latent Models and Bayesian Estimation
 - Longitudinal Data Analysis Techniques
 - General Methodological Issues
 - Adjusting for Non-Response and Attrition
- 최근 주요 이슈를 기반으로 중심이 되는 특정 주제를 정한 다음 이에 따라 keynote address를 구성하고, 전체 세션은 조사방법론 전반에 관한 최신 동향 및 연구과제, 각 국의 현황을 주제로 발표됨

(구성 및 운영)

- 프로그램 구성
 - (개회사 및 폐회사) 개회사는 캐나다 통계청의 방법론국 Assistant Chief Statistician이, 폐회사는 Director General of the Methodology Branch가 함
 - (기조연설 및 수상연설) 기조연설은 해당년의 주제와 맞추어 선정, 수상연설은 Waksberg상 수상강연을 하나씩 구성

- (세션 구성) 기조연설과 수상연설은 통합진행하며 이를 제외하고는 2개 혹은 3개의 세션을 parallel하게 진행, 2104년은 2개 세션을 동시 진행
- (워크숍) 2008년부터 3개씩 시행, 2014년은 시행되지 않음
- (포스터 세션) 최근 시행되지 않다가 2014년에 다시 구성, 커피브레이크를 이용하여 시행
- (토론) 발표별 패널 토론은 없음. 2014년의 경우, 마지막 통합 세션에만 토론 진행

○ 심포지엄 운영

- (발표논문) 양적 발표 논문 수는 캐나다가 제일 많음, 그러나 주요세션은 해외사례가 발표되고 미국 및 호주, 유럽의 발표가 다수 포함됨
- (참석인원) 20개국 500여명 내외, 2014년의 경우 470명 참석
- (등록비) 2014년의 경우, 일반 : CAD 603.62(early-CAD 546.13)
학생 : CAD 287.44(early-CAD 229.95)
- (자료제공) 홈페이지를 통해 목록은 모두 제공, 발표자료 및 논문은 저작권 문제로 홈페이지를 통해 제공하지 않음

2. 2014년 심포지엄 주요 논문

주제

Keynote

□ Survey Sampling in Official Statistics - Some Thoughts in Directions

✍ Ray Chambers, National Institute for Applied Statistics Research Australia(NIASRA), University of Wollongong, Australia

- 최신 이슈는 공식통계에서 추정을 관점을 바꾸는 방법이 필요한가에 대한 논의임
 - 추정방법 변화: design-based → model-assisted → model-based
- 최근 모형을 보조정보로 사용한 추정방법이 시도되고 있지만, 향후에는 모형기반 추정방법을 지향해야함
- 모형기반 추정방법은 베이지안 이론을 이용하면 공식통계에 적용할 수 있음(예: calibrated Bayes)
- 모형기반 추정방법을 적용할 수 있는 분야는 연계자료 분석, 포함오차, 소지역추정, 적응조사설계, 네트워크 모형 등임

주제

Data Collection

□ Effect of using Mobile Devices to Complete the ACS on Data Quality and Respondent Burden

✍ Rachel Horwitz, U.S. Census Bureau, USA

- 2013년 ACS(The American Community Survey)는 인터넷조사방법(개인 PC 또는 모바일)을 추가함
- 태블릿과 스마트폰의 사용 증가로 모바일 조사방법이 선호되고

있지만, 대기시간 증가, 작은 글씨, 선택에 대한 스크롤 증가로 단점들도 많이 부각되고 있음

- 2013년 ACS 조사과정자료를 토대로 컴퓨터, 태블릿, 스마트폰 간의 조사중단빈도, 응답소요시간, 응답변화, 응답오류 등의 자료품질을 비교함
- 전통적으로 응답률이 낮은 집단의 경우는 모바일(태블릿, 스마트폰)에 의한 조사의 응답률이 증가하는 것으로 나타남
 - 이러한 모집단을 추정해야 하는 조사라면 모바일 조사방법을 적극 사용하는 것을 고려해볼 수 있음

주제

Data Collection mode Effect

□ Inference in Surveys with Mixed-mode Data Collection

✎ Jan van den Brakel, Statistics Netherlands and Maastricht University, Netherlands; Bart Buelens, Maastricht University, Netherlands

- 표본조사에서 혼합모드를 이용한 자료수집은 매력적인 방법임
 - 조사비용 및 비표본오차를 줄일 수 있고, 새로운 조사방법을 이끌어낼 수 있는 기술적 개발에 기여할 수 있음
- 혼합모드의 문제점은 주기적으로 조사하는 통계의 경우 응답자의 분포가 모드의 차이에 의해 달라질 수 있다는 점임
 - 그 원인은 측정편향과 다양한 응답소요시간 때문일 것임
- 혼합모드로 주기적으로 조사하는 통계의 시계열은 단일모드로 조사하는 통계와는 달리 편향추정값이 생산될 가능성이 있음
- 혼합모드로 주기적으로 조사하는 통계에서 로버스트한 추정값을 얻기 위한 방법은 다음과 같음
 - 일반화회귀추정량: 설계기반방법으로 조사자료의 관찰값의 측정오차모형을 통해 보정하는 추정량

- 선형모형추정량: 각 모드별로 측정오차와 응답 예측값을 이용한 추정량

주제

Record Linkage

□ Design-based Estimation with Record-Linked Administrative Files

✎ Abel Dasylva, Statistics Canada

- 하나의 행정자료로 밝힐 수 없는 변수들 간의 관련성 등에 대한 정보를 알아보기 위해 여러 행정자료를 활용하는데는 정확매칭이 필수적인 기법임
 - 동일 개체의 연계 키변수가 모두 정의되어 있는지가 중요함
- 행정자료 연계를 통해 얻어진 정보를 토대로 모집단의 평균, 총합, 비율 등 모수를 추정할 수 있음
- 행정자료는 연계키 변수가 유일하지 않을 수도 있고 입력 오류가 발생할 수도 있기 때문에 연계과정은 복잡하고 오류가 발생할 가능성이 높음
 - 그 결과, 연계오류, 적절치 않은 오류(bad linkage), 결측연계 등이 발생하게 되어 편향추정량을 얻게 됨
- 따라서 불편추정량을 얻기 위해 행정자료 연계 후에 방법론 적용이 필요함
- 기본적으로 전통적인 표본조사방식을 이용하되, 설계기반 추정방법을 사용함
 - 매칭된 행정자료에서 확률표본을 반복적으로 추출하여 이를 통해 설계기반 추정량을 획득함
 - 이 방법은 설계기반 표본오차를 가지는 추정량을 얻을 수 있음
- 이러한 방법은 행정자료와 조사자료의 연계결과에도 적용할 수 있음

□ What Big Data May mean for Surveys

☞ Mick P. Couper, survey Research Denter, University of Michigan, USA

- 최근 대규모의 표본조사의 미래에 대한 질문에서 두 가지 현상이 증가하고 있음
 - 첫 번째는 조사비용과 무응답률이 증가함에 따른 효율성과 조사결과의 값어치이고 두 번째는 연속조사가 반복됨에 따른 빅데이터의 관심임
- 빅데이터는 여러 형태로 존재하지만 대부분의 관심은 소셜미디어로부터의 자료수집에 한정되어 있음
- 조사관점에서 빅데이터의 중요성은 고품질의 조사와 빅데이터 분석의 역할을 동시에 수행할 수 있다는 점임
- 빅데이터가 고품질의 조사의 역할을 수행하기 위해서는 모집단 기반의 추론을 적용할 수 있어야함

□ The 2011 National Household Survey Public Use Micro-data File Methodology - "How to balance the requirements for more information and the requirements for low risk of disclosure in the micro-data?"

☞ William Liu and Francois Verret, Statistics Canada

- NHS(The 2011 National Household Survey)는 캐나다 인구총조사의 long-form 질문지를 대신해 실시되는 자발적인 조사임
- NHS의 표본은 캐나다 전체 가구의 약 30%를 조사하고 이중

69%가 응답했음

- 직전 캐나다 인구총조사의 long-form 조사는 전체 가구의 20%를 조사하며, 응답률이 94%였음
- long-form 조사의 마이크로 자료(PUMF; Public Use Micro-data File)는 두 가지 형태로 제공됨
 - 개별 개체 PUMF: 개별 개체의 정보를 모두 포함
 - 계층적 PUMF: 개별 개체 간의 가구·가족 관계에 대한 정보를 제공
- 두 자료가 겹치지 않고 캐나다 전체를 포함하기 위해 부표본 방법을 적용함
 - 이를 만족하면서 자료를 공개하기 위해 자료의 정보량은 높이고 위험도는 낮추는 것이 요구됨

주제

Alternative Sampling Methods

□ Estimation with Nonprobability Surveys and the Question of External Validity

✎ Jill A. Dever, RTI International, USA

- 목표모집단으로부터 알려진 확률추출방법을 사용하는 확률표본 설계는 조사의 주요기준을 고려하여 설계됨
- 1930년대부터 발전해온 표본으로부터 모집단을 추정하는 방법은 design-based, randomization based, model assisted 등이 개발됨
- 전통적인 확률표본은 두 가지 중요키를 가지고 있음
 - 목표모집단과 표본의 연계는 알려져 있고 무응답은 중요한 부분이 아니라는 점임
 - 추출확률의 역수로 가중치를 계산하여 추정을 해야 함
 - 무응답은 랜덤하게 발생하며, 이는 가중치 작성단계에서 조정함
- 현재의 확률표본은 무응답이 증가하고 있고 많은 가정들이 만족

되지 못하므로 비확률추출방법을 고려해야만 함

- 확률표본보다 저비용, 신속성을 만족하는 비확률표본을 활용한 조사는 증가추세이며, 자발적 패널조사 등 다양한 방법으로 시도되고 있음
- 비확률표본을 활용한 추정·분석 등에 대한 연구가 지속적으로 이루어져야 하며, 많은 조건·가정 등을 만족하지 못하는 것에 대한 경고를 제공하여 이용자들이 올바르게 사용할 수 있도록 해야 함

III 시사점

- 표본크기의 감소, 응답률 저하 등을 해결하기 위해 추정방식의 패러다임의 변화를 시도함
 - 설계기반 추정과 모형을 보조정보로 이용한 추정방식에서 모형기반 추정방법으로의 변화를 모색함
 - 모형기반 추정방법에서 매우 중요한 역할을 차지하는 베이지안 추정방법에 대한 지속적인 연구가 수행되고 있음
 - 조사예산의 삭감 등에 대처하기 위한 행정자료 활용에 대한 지속적인 사례연구가 진행되고 있고, 행정자료 연계 등 효율성을 높이는 방법에 대한 논의가 지속되고 있음
- 최신 정보통신기술을 활용한 자료 수집에 대한 연구가 지속적으로 수행됨
 - 표본조사와 같은 전통조사 방식에서 벗어난 디지털 자료수집에 대한 연구가 활발함
 - 태블릿 PC, 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 행정자료, 등록자료에 대한 지속적인 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
- 특히, 행정자료의 활용성을 높이기 위한 방법으로 행정자료 간의 연계방법에 대한 연구가 관심을 받고 있음

□ 심포지엄 준비 및 진행상황 벤치마킹

- 발표진행 및 토론에 충분한 시간을 할애하여 심도 있는 논의가 될 수 있도록 진행됨
- 특히, 질문하는 마이크 자리에 줄을 서서 질문하도록 유도하여 질문 및 토론이 원활하게 진행됨
- 각 세션 중간에 위치한 포스터 세션은 해당 의제를 일목요연하게 정리하여 다양한 내용을 쉽게 접할 수 있는 기회가 됨
- 회의 일정 및 발표자료의 단락요약에 대한 간단한 팸플릿만 제공됐을 뿐 발표자료나 논문 등을 축약한 책자가 없어 발표 내용을 전체적으로 파악하는 데 다소 어려움
- 발표자료의 저작권 문제로 발표자료 전체가 인터넷, USB 등 어떠한 형태로도 제공되지 않아 향후 활용 및 적용 면에서 효율적이지 않음