

2009년 FCSM 컨퍼런스 참석 결과

1 출장 개요

- 출장명 : 2009 Federal Committee on Statistical Methodology Research Conference(미연방정부 통계방법론 연구회의) 참석
- 기 간 : 2009. 11. 1. ~ 11. 6.(4박 6일)
- 출장지 : 미국 워싱턴 D.C.
- 출장자 : 표본관리과 5급 김진, 7급 김세미
- 목 적 :
 - 미 연방 정부에서 실시하는 각종 통계조사 표본설계 방법론 및 분석기법을 습득하여, 우리나라 현실에 맞는 표본설계 기법의 개발 및 정착화 방안을 모색
- 참가 내용
 - 국가범죄피해조사(NCVS)의 재설계
 - 행정자료의 통계적 이용
 - 포함률과 응답률의 개선
 - 표본설계와 추정의 개선
 - 농업 표본추출틀 개발 및 사용
 - 분산추정
 - 조사 추정에서의 효과적인 대체
 - 조사 오차 감소 원인
 - 회수 및 응답에 대한 조사표와 표본설계의 효과

1. 컨퍼런스 소개

○ FCSM 컨퍼런스란

- FCSM은 Federal Committee on Statistical Methodology의 약어로 미연방의 국내 통계기관들이 생산하는 통계의 질을 향상시키고, 통계기관들에게 통계적 방법론의 기준을 제안하기 위해 현업의 통계학자, 경제학자, 통계기관 관리들이 모이는 회의
- 최초의 회의는 1975년 여름 개최되었고 이후 매년 개최되고 있음(다음 회의는 2011년 11월 개최)
- 공식 웹사이트는 www.fcsm.gov이며 여기서 지난 회의의 주제도 찾아볼 수 있음

○ 컨퍼런스 개최 목적

- 모든 미연방 통계기관들 간에 현업의 정보를 교환하고 보급함
- 미연방 통계 프로그램에서 생산되는 자료의 질을 높이기 위한 새로운 방법론을 소개함
- 서로 다른 통계기관 종사자 및 전문가들이 만나서 의견을 교환하는 자리를 제공함

○ 컨퍼런스 분위기

- 통계기관 근무 경력의 길고 짧음에 상관없이 각종 기관들의 현 업무 담당자 및 통계 전문가들이 갖고 있는 다양한 주제들이 한번에 세 가지 세션으로 진행되었으며 15~20분 내외로 발표하고 토론
- 주제는 학술적인 것 보다는 업무에서 활용하고 있는 실질적이며 기술적인 방법론이었으며 한국에서는 경험적으로 처리하여 넘겨질 문제조차 진지하게 발표하고 의견을 나누는 분위기이었음
- 참가 주체는 미국이었으나 호주, 캐나다, 프랑스, 이탈리아, 싱가포르, 네덜란드, 베트남에서도 참여함

2. 주요 주제발표 요약

1) 국가범죄피해조사(NCVS)의 재설계

- Assessing Sample Design Options for the National Crime Victimization Survey
 - Robert Fay and Jianzhu Li, Westat, USA
 - 표본수가 1972년 72,000가구에서 2005년 38,600가구로 감소
 - .. 자료수집 비용 증가대비 예산은 감소하여 점차적으로 표본수를 줄임
 - .. 애초의 목표에 도달하지 못하는 결과를 가져옴
 - 대안을 제시
 - .. 가구조사 설계에서 subsampling하는 것 고려
 - .. 영국범죄조사(British Crime Survey)에서와 같은 패널조사를 고려해봄

- Analysis of Possible Nonresponse Bias in the NCVS
 - Fritz Scheuren and Zhiwei Zhang, NORC at the Univ. of Chicago, USA
 - NCVS는 응답률이 90%이상이지만 미국내 다른 가구 대상 조사의 응답률이 감소추세이므로 무응답의 증가로 인한 주요 결과는 조사비용 증가와 잠재적 편향을 가져옴
 - 다음의 네 가지 방법으로 무응답 편향을 분석하였고 이것은 OMB 가이드라인에 의한 것임
 - .. 2개 이상의 Wave에서 평균응답성향점수를 얻기 위한 Capture/Recapture 모델 접근법
 - .. 조사 결과를 쉽게 얻는 응답자와 어렵게 얻어내는 응답자의 차이를 비교
 - .. 전화 기록 Paradata와 log를 사용하여 주요 부차그룹내의 응답률과 특성을 비교
 - .. 표본추출틀의 변수 혹은 자료와 결합될 수 있는 외부 자료를 이용한 응답자와 무응답자들 간의 비교

※ Paradata

- 조사행위의 전반에 걸쳐 다양하게 수집되는 자료로 “process data”라고도 함
- 면접자 각종 경력, 면접성공률, PSU특성, 층정보, 해당주소가 선택될 확률, 총 전화통화 시도 횟수, 최초 통화상태, 전화 건 날짜, 시간, 최종상태, 자료제공자, 접촉결과 등
- 조사환경을 정확히 이해하고 조사원의 활동이나 조사방법의 흐름을 모니터링하기 위해 수집하는 자료로써 조사설계, 조사관리, 비용대비 효과를 측정하는데 유용하게 사용

○ Evaluation of a 12-Month Reference Period in the NCVS

- Lisa Lee and James Carr, NORC at the Univ. of Chicago, USA
- 조사기준 시점의 길이가 6개월에서 12개월로 길어짐에 따라 비용면에서는 유리해졌으나 자료의 질은 낮아질 위험이 있음
- 사건 회상의 정확성 향상과 응답자 부담 경감을 위해 두 가지 방법을 개발하고 테스트
 - .. 기억력 향상에 도움을 줄 수 있도록 사건기록 일지활용(EHC) 및 향상된 사건 전후관계 주입법(ECP)
 - .. 시카고 지역에서 전화조사와 인터넷조사로 시험조사

2) 행정자료의 통계적 이용

○ An Evaluation of Data from the NCES Teacher Compensation Survey :

The Relationship Between Teacher Compensation and Students in Poverty

- Stephen Cornman and Frank Johnson(U.S. Department of Education) et al.
- National Center for Education Statistics(NCES)와 Teacher Compensation Survey(TCS)의 결합 시도
- FCSM 발표 목적 중 하나는 추가 참여를 독려하기 위함
 - .. 2007년 TCS 시험자료 수집에는 7개주 참여
 - .. 2008년 TCS 시험자료 수집에는 17개주 참여(전국 1/3수준)
- 자료 수집의 방법론과 문제점을 공유하고자 함
 - .. 17개주 중 5개 주만이 건강과 퇴직비용 자료를 보유함
 - .. 교사경험, 학력, 연령, 인종, 성별 교사의 월급 중위수 소개

○ Using Administrative Data in the 2007 Census of Agriculture Edit

- Virginia Harris et al.(National Agricultural Statistics Service, USA)
- NASS의 농업 센서스 자료와 Farm Service Agency(FSA)의 자료
 - .. 5년 주기의 농업 센서스에서 미농무부의 농가 프로그램 보상에 대한 정보를 수집하고 있음
 - .. 과거 경험에 비추어 농가 프로그램 보상은 센서스에서 적게 보고됨
- 2007년 센서스의 총계를 개선하기 위해 행정자료 사용 시도
 - .. FSA의 농가 프로그램 보상 자료와 결합

- FSA의 보상자료와 NASS의 농가보상유무자료를 결합하는데 어려움이 있었지만 공표된 총계는 FSA에서 지불한 총계와 비슷해짐
- 향후 계획은 좀 더 정교하고 강건한 결합 방법을 사용하여 공표수치를 개선시키는 것과 행정자료 결합을 이용하여 센서스 응답 부담을 경감시키는 것

3) 포함률과 응답률의 개선

○ SIPP 2008 Incentive Analysis

- Stephen Clark and Stephen Mack(U.S. Census Bureau)
- Survey of Income and Program Participation(SIPP)의 응답률을 제고하기 위한 효과적인 인센티브 전략 마련
 - 인센티브 제공이 응답률을 높이는지와 어떤 식으로 제공해야 가장 좋을지 연구하기 위해 2008년 SIPP 패널 1차, 2차 Wave를 대상으로 임의 실험
 - 패널 중 50%는 인센티브 제공(조사 전에 \$20, 조사 참여자에 한하여 응답 완료 후 \$40제공 \$40제공을 약속) 나머지 50% 미제공함
- 2008년 9월~12월의 기간 동안 실험결과
 - 인센티브 제공유무에 따른 응답률과 그에 따른 오차, 가구거절률과 그에 따른 오차를 분석
 - 12개 지역 사무소별로 응답률을 분석한 결과 미제공그룹과 제공그룹간의 차이가 크지 않음. 단 2개 지역(캔자스시티와 시애틀)의 경우 차이가 남
 - 제공하는 편이 응답률이 약간 상승하였고 \$20와 \$40의 차이는 없었음
 - 수입이 적은 그룹과 큰 그룹간에도 인센티브의 차이가 유의한 응답률의 차이를 나타내지 않음

○ Adaptive Contact Strategies in a Telephone Survey

- James Wagner(Univ. of Michigan, USA)
- 가구대상의 전화조사에서 적절한 통화시간대를 정하는 것이 중요
 - 일반적으로 최초통화에 평균 최적 통화시각이 사용됨
 - 그러나 다음번 통화시각은 가구의 특성별로 맞춰져야 할 필요성이 있음
- Multi-level 모형을 통해 최적의 통화시도 시간대를 추정하여 제시
 - 이 모형은 다음번 통화시간대의 결정을 위해 일일간격으로 갱신됨

- Characteristics of Parent Respondents in the National Household Education Surveys
 - Stacey Bielick and Cameron Mcphee(American Institute for Research, USA)
 - NCES에서 주관하는 NHES의 응답자(부모)의 인구통계학적 특성을 분석
 - .. NHES는 1991년 최초 조사되었으며 2007년이 가장 최근의 조사임
 - .. 전화조사로 행해지며 점차적으로 응답률이 감소추세
 - .. Screener 응답률이 1991년 80%에서 2007년 53%로 감소
 - 응답자 특성에 의해 주요 관심 추정치가 달라짐
 - .. 결과를 바탕으로 전화조사에서 웹조사로 바뀌는 추후 설계에 반영할 예정
 - .. 응답자의 대부분은 여성(어머니)이고 최근에 남성(아버지)이 증가추세
 - .. 부차모집단(아이의 연령 및 교육방법에 따른 분류)에 따라 응답자(부모)의 특성이 다름(홈스쿨을 하는 응답자가 특별히 구분되는 성질을 가짐)
 - .. 스페인어 사용자가 증가, 모집단 경향보다 과도하게 고학력 응답자의 초과근무 증가, 빈곤상태의 변화

- Measuring the Effects of Reminder Postcards in the Occupational Employment Statistics Survey
 - Carrie Jones(Bureau of Labor Statistics, USA)
 - 직업별 고용과 임금추정을 위한 조사로 4주간의 간격을 두고 3번 우편조사
 - .. 120만 기업이 3년간 패널로 응답
 - 독촉우편엽서의 효과 분석 배경 및 방법
 - .. 2007년 우편발송비용이 30%인상 되었으나 예산은 축소
 - .. 많은 응답자들이 조사를 완료하여 회신한 이후에 조사권유 패키지를 받는 것으로 나타남
 - .. 조사권유 패키지를 대신하여 2007년 11월, 2008년 5월에 독촉우편엽서를 보내기로 결정하고 6개 주에서 우편엽서를 받는 그룹과 기존의 방식대로 패키지를 받는 그룹으로 나누어 실험
 - .. 응답률에 있어 효과적인지를 로지스틱 회귀분석으로 분석하고 비용분석 및 각 주별 질문지의 질적인 완료를 평가함
 - 결과
 - .. 비용감소는 효과적이며 응답률에는 약간 영향이 있음
 - .. 우편엽서를 받은 중소기업은 응답률이 약간 낮아진 반면 대기업은 증가

- .. 중소기업에는 우편엽서를 보내고 대기업은 완전한 조사패키지를 보내는 것이 비용 대비 효율적임
- .. 중소기업은 우편엽서, 대기업은 우편엽서와 일반 조사패키지를 보내는 것이 확실한 비용절감과 응답률 감소 최소효과를 나타냄

4) 표본설계와 추정의 개선

- An Evaluation of the Use of the ACS as the Sampling Frame for the NIS
 - James Singleton and Philip Smith(Center for Disease Control and Prevention, USA) and Andrea Piani et al.(U.S Census Bureau)
 - NIS에서 ACS 조사자료를 표본추출틀로 사용이 가능한지 검토
 - .. 현재 NIS는 전화로 하는 적격자 심사 조사에서 적격대상인 19~35개월 유아를 돌보는 가구를 3~4%정도 확인
 - .. ACS는 NIS의 조사모집단을 전부 포함하지 않고 있다는 제약이 있지만 현재 사용하는 표본추출틀보다는 나은 정보를 제공
 - .. 유선전화 없이 휴대폰만 소유한 가구도 포함하기 때문
 - .. ACS에 참여한 19~35개월 유아를 돌보는 가구를 검토대상으로 함
 - NIS의 자료수집방법
 - .. 1단계 : 전화조사로 적격가구인지 그리고 부모가 백신공급기관 자료와 연결을 허락하는지를 걸러냄
 - .. 2단계 : 백신공급기관에서 백신접종상태를 우편조사
 - - .. 휴대폰만 소유한 가구나 전화조사 방법을 거부하는 가구도 조사가 가능
 - .. 적격 가구에 대한 조사 포함률을 높임
- Methodological Changes in the Job Openings and Labor Turnover Survey
 - John wohlford(Bureau of Labor Statistics, USA)
 - JOLTS란
 - .. JOLTS는 미국 직업 시장의 노동자 수요, 고용, 이직의 정보를 제공
 - .. 공적 사적부문의 노동자들로부터 월별 취직자리, 고용, 이직을 조사
 - CES 고용동향과의 차이
 - .. JOLTS의 고용과 이직에 대한 자료가 더 크고 잘 알려진 CES조사 만큼

잘 맞지 않음

- BLS는 CES의 월별 고용동향과 JOLTS 고용의 차이를 밝혀내기 위해 다년간 연구
- 연구결과
 - 표본설계를 개선, 출생/사망 모형 도입, 마이크로 자료 재검토 과정을 통한 수정, CES고용동향을 입력자료로 사용한 조정
 - 모든 JOLTS 시계열을 재계산 및 재공표

5) 분산추정

- The RAO-WU Rescaling Bootstrap: From Theory to Practice
 - Claude Girard(Statistics Canada)
 - RAO-WU 부스트랩
 - 캐나다 통계청에서 분산계산할 때 가장 많이 쓰는 방법
 - 에프론의 원래 방법을 표본조사에서의 분산계산에 적용한 방법
 - 가중치를(부스트랩 가중치) 증가시키는 방법론과 쉬운 사용법
 - RAO-WU 부스트랩 사용에 따른 문제점 발생 및 해결 방안 검토
 - RAO-WU 부스트랩의 사용이 증가하면서 이상한 분산추정치 등장하여 이에 대한 방법론 검토
 - 5가지 Case Study 제안

- Linearization Variance Estimation and Allocation for Two-Phase Sampling Under Mass Imputation
 - Abdellatif Demnati(Statistics Canada) and J.N.K. Rao(Carleton Univ.,Canada)
 - 관심변수의 값이 두 번째 단계에서 관찰되는 이단계 표본추출
 - 첫 번째 단계에서 추출되지 않은 단위는 두 번째 단계에서 행정자료를 이용하여 집단 대체되거나 회귀대체 됨
 - 이단계 표본추출은 비용절감과 응답자 부담경감 등의 이유로 연간 사업체 조사에서 종종 사용됨
 - 행정자료비용이 조사자료 비용보다 훨씬 저렴하다는 전제 하에 사용
 - 미경험적이고 설계와 일치된 총합 또는 평균 추정량 연구
 - 유한모집단에서 추정량과 분산추정량의 성능을 시뮬레이션

- On Point Estimation and Variance Estimation for Mean Expenditures in the Consumer Expenditure Survey
 - Michail Sverchkov, John Eltinge et al.(Bureau of Labor Statistics, USA)
 - CE조사는 다단설계
 - .. PSU(지역)를 추출 후 예산에 의해 최종 PSU의 수를 조정
 - .. 평균 지출은 Hajek방식에 의해 비추정
 - .. 이때 23개의 알려진 인구통계학적 모집단 총합을 이용해 가중치를 사후 조정
 - .. 분산은 BRR방법으로 계산
 - .. PSU 표본수 조정이 현재의 평균과 분산추정에 어느 정도 영향을 주는지 검토

6) 농업 표본추출틀 개발 및 사용

- Utilizing An Alternative Sampling Frame to Produce Agricultural Survey Indications
 - Wendy Barboza and Mark Harris(National Agricultural Statistics Service, USA)
 - 미국 농무부의 NASS는 매년 수백 개의 조사를 수행하고 보고서를 작성
 - .. 표본설계에서 리스트 추출틀과 지역 추출틀을 조합하여 사용하였음
 - 응답부담 경감, 자료수집 비용절감 등을 위한 표본추출틀 개발 노력
 - .. 생산자가 매년 작물재배면적을 보고한 행정자료를 사용하기 위해 USDA의 Farm Service Agency(FSA)와 수년간 함께 일해 옴
 - .. 2006년 12월부터 FSA자료를 표본추출틀로 사용하기 위한 검토작업을 2년간 수행
 - .. FSA자료가 표본추출틀로 부적합하다는 결론을 내렸으나 FSA 행정자료의 다른 사용방안을 계속 찾고 있음
 - .. FSA 행정자료 소개 및 위 사례를 연구한 방법 설명함
- Four Digits or No-Digit Social Security Numbers-Impact on the National Agricultural Statistics Service Record Linkage Maintenance Processes
 - Denise Abreu et al.(NASS, USA)

- NASS는 표본추출명부를 행정자료와 연결하여 사용
 - .. 농장주와 목장주에 대한 양질의 명부를 확보하기 위해 사회보장번호(SSNs)와 고용주등록번호(EINs)를 연결하여 사용하여 왔음
 - .. NASS에서 구별해 낼 수 없었던 새로운 농장주와 목장주를 구별해 낼 수 있고 그 외의 정비목적(명부의 전화번호 갱신 같은)을 위해 자료 연결
 - .. 명부상의 9자리 SSNs/EINs를 유지하는 것에 문제가 발생
- SSNs/EINs 정보를 명부상에서 제거하는 방안과 4자리의 숫자만 남겨놓는 방안을 검토
 - .. 4자리 숫자를 남겨놓는 경우 자료 연결시 SSN은 1~2%, EIN은 4~6%의 미결합 자료 발생
 - .. 완전히 제거 하는 경우 SSN은 4~6%, EIN은 9~13%의 미결합 자료 발생
 - .. 다양한 자료 연결을 시도하였고, 잘못된 연결은 연결시키는 자료의 크기와 관련됨

○ Comparison of Methods for Updating Census Based Estimates of Number of Farms to Non-Census year

- Michael Bellow(NASS) and Phillip Kott(RTI International, USA)
- 농업센서스가 미시행되는 해의 농장수의 추정에 대한 연구
 - .. 농업센서스는 5년마다 시행
 - .. 연간 표본조사의 추정값을 활용하여 센서스가 미시행되는 해의 특정 농업 물품의 추계의 정확도를 높일 수 있음
- 이를 위해 두 가지 방법을 제안
 - .. June Area Survey(JSA)를 사용하여 현재 년의 총계를 만들어냄
 - .. JSA와 함께 NASS가 공표한 전년도 주별 추정 농장수를 추가로 사용
 - .. 두 가지 방법에 의한 결과물을 비교
 - .. 잭나이프(한 그룹씩 제거 하는) 분산추정 사용
- 결론 요약
 - .. 두 방법이 분산추정과 절대오차가 작다는 측면에서 뛰어난 성능을 보임
 - .. JSA와 함께 NASS가 공표한 전년도 주별 추정 농장수를 추가로 사용한 두 번째 방법이 좀 더 나은 결과를 나타냄

3. 시사점 및 결언

○ FSCM 컨퍼런스 특징

- 분산형 통계체제인 미국은 각 기관이 생산하는 국가통계를 개선하기 위해 다양한 방법 및 관점으로 설계를 기획하고 있으며 방법론의 공유가 활발히 이루어지고 있음
- 사소하게 느껴지는 조사 방법론의 차이도 개편기간 동안 시험 조사를 통해 효과를 확인해 본 후에 현실에 적용하려고 함

○ 한국 조사환경과의 차별성

- 전화와 인터넷 사용추세가 강함(1단계 우편, 2단계는 1단계 미응답자 대상으로 전화조사, 3단계는 2단계 무응답자를 대상으로 면접조사)
- 이러한 상황에서 조사 및 표본설계에 대한 시각이 한국과는 다름을 알 수 있었음(한국은 미국과는 달리 지리적으로 넓지 않기 때문에 현재의 면접조사 유리할 것으로 판단됨)
- 행정자료의 통계적 이용, 응답자 부담경감과 비용감소를 위한 표본추출틀 사용 전략, 정확한 추정을 위한 조사 및 표본설계 계획 등을 참고하여 향후 업무에 활용

【 참고 】 세션별 회의주제 및 발표논문 초록(별첨)