PPSM
Priority Programme on
Survey Methodology

제2차 국제 방법론 학술회의 참가 결과보고

2nd International Conference in Bremen
"Advancing Survey Methods"

2011. 12.

미래로 통하는 빛
통계청

Ⅰ 개요

1. 출장 개요

◦ 출장 목적

- 선진 주요국 통계청을 포함한 학계 및 서베이 관련 민간기관은 조사환경의 악화에 대응한 통계작성의 효율화 및 정확성 제고를 위해 서베이 방법론에 대한 연구 경험을 공유하는 한편 개선방안을 활발히 모색중에 있음

- 이번 국제 방법론 회의도 최근의 서베이 방법론의 주요 분야에 대한 최근 연구성과를 집약적으로 전달하고, 토론하는 장임

- 본 출장자들은 서베이 방법론 개발의 차원에서 관련 선진 통계기법을 파악하고, 이를 실용적으로 적용하기 위한 연구를 적극적으로 추진하고자 하는 바, 이번 회의 참석과 협의를 통해 소기의 목적을 얻고자 함

◦ 참가회의

- 제2차 국제 방법론 학술회의(2nd International Conference in Bremen)

◦ 출장자

- 표본과 김규영 과장, 조사연구실 한성호 실장

◦ 출장기간

- 한성호 : 2011. 11. 15.(화) - 11. 20.(일)

- 김규영 : 2011. 11. 16.(수) - 11. 20.(일)

◦ 출장지

- 독일 브레멘

2. 회의 개요

◦ 회의명

- 선진 서베이 방법론(Advancing Survey Methods)

◦ 회의일시 및 장소

- 2011. 11. 17. ~ 11. 18. (2일간)

- 독일 브레멘 대학

◦ 참가자

- 약 70명 서베이 방법론 전문가

- 독일, 스위스, 네덜란드, 영국, 남아공, 슬로베니아 연구자

◦ 주요 업무 내용

- 회의 참석 토의

	토의 주제
1일차	Nonresponse
	Sensitive Questions
	Web Surveys
	Factoral Surveys
2일차	Interviewer and Measurement Error
	Access Panels
	Sampling
	Linking Survey Data

- PPSM에 대한 업무협의

• 이 회의와 연관된 "서베이 방법론 중점연구 프로그램(PPSM: Priority Programme on Survey Methodology)“은 2008년 5월에서 2012년 4월까지 672만 유로가 투자되어 독일, 연국, 네덜란드의 학계 네트워크를 중심으로 추진되어 왔고, 향후 연장될 계획임

• 이 프로그램의 추진 중심인 브레멘 대학 엥겔(Ulrich Engel) 교수를 방문하여, 본 프로그램의 경험을 벤치마킹하는 한편 관련 분야 공동연구 등 협력 가능성을 협의

II 회의 내용

주제1 Nonresponse

❒ Nonresponse in a cross-national survey. Enhancing response rates and using auxiliary data to detect and adjust for nonresponse bias

✍ Ineke Stoop

◦ 유럽사회조사(ESS: European Social Survey)는 가치, 태도, 의견, 신념에 대해 30개국 이상에서 2년마다 이루어지는 면접조사이고, 방법론상의 품질 향상을 지향

◦ ESS는 70% 이상의 응답률을 목표로 설정하고 있고, 조사대상 국가간에 서베이의 설계와 실행이 가능한 한 동질적인 환경에서 이루어지도록 함

◦ 여러 노력에도 불구하고 70% 목표가 이루어지고 있지 못한 나라가 많고 응답률은 국가간에 또한 시기별로 다름. 이러한 차이의 원인으로 여러 요인들이 확인되지만 분명치 않음

◦ 몇몇 연구를 통해 응답률과 무응답 편향간에는 직접적인 관계가 없다는 것이 밝혀져 왔고, 무응답 편향의 존재와 크기는 활용 정보가 많지 않은 가치와 태도 조사에서 평가가 어려움

◦ ESS에서는 개인통화, 접촉 결과 및 거절 이유 등의 접촉정보가 자세히 기록되고, 여러 종류의 부가정보와 패러데이터가 무응답 편향을 확인하기 위해 사용되어 옴

❒ Costs of increasing survey response – Experimenting with monetary incentives

✍ Ulrich Krieger and Mathis Schröder

◦ 기존 면접조사에서 응답률을 증가시키거나 혹은 현재의 수준에서 유지시키려는 시도로 서베이 연구자들은 금전적 인센티브 방안을 고려

◦ 서베이에 대한 협력으로 응답자에게 금전을 주는 것은 응답자가 들인 시간과 노력을 보상하고 그에 대해 고맙다는 신호를 보내기 위한 타당하고, 관리하기 쉬운 방법으로 여겨지나 응답자의 반발도 발생

◦ 독일의 경우 금전 인센티브의 효과는 어떠할 것인가를 아래와 같은 문제의식하에 연구

- 현금 제공이 응답자의 순수한 동기를 방해하고, 금전적 거래의 좋지 않은 형태로 인식되지 않을까?

- 이러한 인센티브의 순효과는 어떨까? 그것이 너무 높은 수준이어서 응답자들에게 뇌물로 인식되어 응답을 기피하게 할 것인가 아니면 많이 줄수록 많이 응답할 것인가?

- 금전 인센티브는 응답자들에게 선택적으로 서베이에 참여하게 함으로써 의도치 않은 편향을 초래하는 것은 아닐까?

◦ DFG(독일학술재단)의 자금 지원을 받아 '유럽 건강, 연령, 은퇴 서베이(SHARE: Survey of Health, Aging and Retirement in Europe)' 독일 재조사 표본을 대상으로 실험

◦ 일정한 금액(0, 10, 20, 40유로)을 대상 표본에게 실험참여 여부 응답과 관계없이 모두 보내 금전 인센티브의 효과를 파악하고, 응답자들에 대해 금전 수준의 편향 발생 여부를 판단

❒ Addressing panel attrition in mobile phone panel surveys: Can incentives and switching to mobile web help? Results from a German mobile phone panel

✍ Britta Busse and Marek Fuchs

◦ 모바일폰 가입자의 증가에 따른 유선전화 가입률의 감소는 모든 연구자들에게 기존 전화 방식의 전환 검토 요구

◦ 핀란드, 스웨덴, 체코 같은 몇몇 유럽 국가들은 모바일폰 가입률이 90%를 넘으므로 모바일폰 서베이가 바람직하고, 또한 패널 서베이 방법을 이용하여 보다 긴 시간동안 추적 필요

◦ 그러나 패널서베이는 패널 표본 이탈 가능성이 높아 데이터 품질이 표본 크기의 감소와 표본 이탈에 의한 편의에 의해 위협받음

◦ 모바일폰 패널의 표본이탈의 이유로는 (1) 패널기간동안 전화번호가 유용하지 않게 되거나, (2) 접촉이 안되는 경우, (3) 더 이상 패널에 참여하지 않겠다고 것을 들 수 있는데, 이를 해결하기 위한 실용적 방법 개발 필요

◦ 이 논문에서는 2009년부터 독일에서 1,700명의 표본규모로 실험적으로 이루어진 모바일폰 패널에서의 표본이탈을 해결하기 위한 2가지 방법의 결과를 제시

- 표본이탈로 인한 이탈률과 편의 수준을 인센티브를 제공받지 않은 통제집단과 비교

- 모바일 웹에 의한 텍스트 입력 방식에 따른 거부과 비접촉의 수준을 비교

주제2 Sensitive Questions

❒ Sensitive questions in surveys using the randomized response- and the item count-technique. Results from an experimental validation survey

✍ Felix Wolter, Peter Preisendörfer and Anna Skarbek-Kozietulska

◦ 민감사항에 대한 서베이(예를 들면, 마약복용, 음주량, 범죄사실 등에 대한 질문)는 응답자의 오류응답 가능성 높음

- 사람들은 바람직하지 않거나 부정적으로 인식되는 사항에 대해서는 작게 응답하고, 바람직한 것에 대해서는 크게 응답하여 민감항목에 대한 추정치는 편의 발생

◦ 민감사항의 서베이에서의 응답오류의 문제를 해결하기 위해 임의응답방식(RRT : Randomized Reponse Technique)과 항목계산방식(ICT : Item Count Technique) 제안되어 옴

- 두 방법의 기본 아이디어는 익명 상태를 조성하여 응답자가 잘못 응답하려는 유인을 제한. 그러나 직접질문(DQ : Direct Question)과 비교하여 우수한가에 대해 상반된 결과 존재

◦ 이 논문은 552명을 표본으로 범죄사실에 대한 몇몇 민감질문(도둑질, 마약복용 등)을 포함하는 시험조사의 결과 제시

- 이 연구의 특징은 응답자들이 실제로 범죄를 저질렀는지에 대한 정확한 정보를 행정기록을 통해 알고 있다는 점

- 세 방식의 응답타당성을 비교하였고, 사실로 응답하지 않는 결정요인을 분석. 실험결과는 DQ와 RRT에서 응답오류가 상당히 발생했음을 보여줌

❒ Using Benford's Law to improve the efficiency of the randomized response technique: An application to student cheating

✍ Andreas Diekmann and Marc Höglinger

◦ RRT는 응답과정에 확률성을 도입함으로써 민감한 질문에 대한 응답자의 내용을 보호

- 주사위 같은 확률 추출 도구는 특정 응답자가 민감 질문에 대해 대답해야 할 것인지의 여부를 고정 확률로 결정

- RRT 실행의 중요한 과제는 민감질문에 대답할 확률의 선택 문제. 통계적 관점에서는 가능한 한 정확한 응답을 얻을 수 있는 확률이 선택되어야 하고, 개인정보 보호의 관점에서는 응답해야 할 확률이 가능한 한 낮은 것이 좋음

◦ 통계적 효율성과 응답자 개인정보 보호의 상충관계를 줄이기 위해 벤포드 분포 수자를 확률 선택의 수단으로 활용

- 벤포드 분포를 따르는 주소의 첫 번째 수자의 사용을 제안

※ 벤포드 분포 : 일반적으로 사람들 주소의 수자 첫 자의 분포는 균등 분포가 아닌 것을 보여줌

- 응답자들은 그 수자가 균등분포를 한다고 생각하고 있기 때문에 낮은 수자가 걸릴 확률에 대해 실제보다 더 적게 생각

◦ 학생들의 부정행위에 대한 온라인 시험조사에서 벤포드 분포에 기초한 확률 선정을 적용한 결과를 다른 RRT나 DQ 방식의 실행과 비교

◦ 그 외에도 민감한 질문에 대해 더 높은 응답률을 제공하도록 선택한 것이 응답자의 개인정보 보호에 대해 혹은 시험에서 부정행위 자백 응답에 미치는 효과에 대해 분석

주제3 Web Surveys

❒ Web survey process within the concept of eSocial Sciences

✍ Vasja Vehovar

◦ 15년간 이루어진 문헌 검토를 통해 웹서베이 방법론의 전반적 추세를 살펴보고, 또한 최근의 주요 연구과제에 대해 제시

◦ 웹서베이를 e-사회과학이라는 보다 넓은 개념틀로 포괄하는 것이 중요

◦ 웹서베이의 분석과 공표에 이르는 전체 연구 과정은 ICT의 지원을 받아 통합되어야 할 것임

◦ 서베이 방식의 전환 현상과 요인에 대해 특별한 관심을 기울여 왔는데, 제한된 부분만이 온라인으로 진행

- 실제 웹서베이 과정으로의 전환에 대한 계량화 수준이 파악되고 토의되어야 할 것임

◦ 완전한 웹 형식을 통해 설문을 개발하는 과정의 실험 결과가 기존의 일반적인 과정과 비교되었음

- 종래의 과정에서는 설문이 (1) 워드문서 양식으로 작성되고, (2) 설문지들의 여러 버전이 이메일이나 온라인 토론 시스템을 통해 교환 토의되고, (3) 다양한 형태의 버전(웹사이트, 공유, 구글문서) 제공을 위해 사용됨

❒ Equivalence of Measures of Xenophobia in Cross-national Survey Research: Findings based on a Web Probing Study

✍ Dorothée Behr, Michael Braun, Lars Kaczmirek, Wolfgang Bandilla and Sandra Majer

◦ 조사간 동등성 확보는 국가가 서베이 자료를 실제로 분석할 때 갖춰져야 할 조건

◦ 인지적 인터뷰는 서베이 자료에서 비동등성의 존재 특히 그 원인을 식별하고자 할 때 적절한 방식

- 통상적인 면접 인지 인터뷰는 표본규모 제약, 시간, 비용 혹은 조사원 효과 등과 같은 약점이나 과제 존재. 특히 국가간 비교 연구에서 인지 인터뷰 방식의 사용은 문제가 있음

◦ 국가간 문화 차이를 연구하고 측정의 동등성을 평가하기 위한 보완적 수단으로서 웹서베이를 검토

- 웹서베이는 통상적 면접 인지 인터뷰의 문제들을 해결하는 역할 수행

◦ 이 논문은 캐나다, 덴마크, 독일, 헝가리, 스페인, 미국에서 2011년 초 실행한 웹서베이에서 동등성의 발견 결과 제시

- 각국에서의 480명씩이 목표 표본규모로 이주자에 대한 시각과 관련된 4개의 항목이 질문으로 포함됨

- 즉, 이주자가 범죄율을 증가시켰는가, 경제에 유익한가, 직업을 뺏아간 것은 아닌지, 사회에 새로운 생각과 문화를 가져다 주었는지에 대한 질문

- 웹서베이에서 응답 행동에 디자인이 차지하는 영향을 중요하게 생각하여 추가적으로 디자인 실험도 시행

주제4 Factoral Surveys

❒ First equals most important? Order effects in vignette-based measurement

✍ Katrin Auspurg and Annette Jäckle

◦ 사람들의 태도, 개념, 결정을 좌우하는 요인이 무엇인가를 측정하기 위해 사회과학 서베이는 응답문구(vignettes) 방식의 사용 증가

◦ 응답문구는 보통 상황이나 대상의 다양한 특성을 묘사. 응답자들에게는 응답문구에 표현된 가설적 상황이나 대상을 보고 판단하도록 질문이 이루어짐

◦ 이전에 수행된 관련 연구는 대상자에 대해 항목이 제시된 순서가 응답자의 판단에 영향을 준다는 결과를 보여줌

◦ 순서효과를 파악하기 위해 인지심리학과 서베이 방법론의 문헌에 기초하여 2개의 주요 연구질문에 대해 실험

- 항목이 제시되는 순서가 응답문구 평가에 영향을 주어 실제 결론을 변화시키는가? 어느 조건하에서 순서효과가 발생할 가능성이 높은가?

- 응답문구 항목의 특성(항목의 수, 평가 업무의 복잡성, 응답문구의 내용, 항목 중요성)과 응답자의 특성(주제지식, 태도의 강렬도, 인지능력), 두 특성들간의 상호작용 등을 분석 한 결과 강한 순서효과가 발생

주제5 Interviewer and Measurement Error

❒ Interviewers, nonresponse, and measurement error

✍ Patrick Sturgis

◦ 조사원의 태도, 인성, 행동이 표본가구와의 접촉과 협력의 성공과 관련되어 있다는 연구가 많이 제시되어 왔음

◦ 그러나 조사원들의 특성이 표준화된 인터뷰를 실행할 때 최선의 방식을 따르는 정도를 결정하는 원인이란 점에 대해서는 별로 잘 다루지 않았음

- 조사원에 의해 유발된 무응답과 측정오차간에는 상관관계가 있다는 것을 실증하기 위한 2가지 연계된 연구 수행

◦ 하나는 응답원의 접촉과 협력을 얻어내는 기술과 관련된 조사원 특성 차이간의 관계의 방향과 정도를 살펴 봄

- 횡단면 자료를 가지고 복합에러 구조를 갖는 다수준 모델을 이용하여 조사원 특성의 차이가 접촉과 협력을 얻는 것과 상관관계가 있다는 것을 검토

- 이를 통해 여러 영역의 변수에 대해 과거 접촉과 협력 실적이 나빴던 조사원의 경우 실적이 좋았던 동료에 비해 상당한 수준의 조사원 오차를 발생시켰다는 사실을 확인

◦ 다른 하나는 조사원에 대해 조사된 데이터를 서베이 결과와 연결시켜 조사원 태도와 신념이 그들이 조사한 응답내용의 측정오차 정도와 연관되는지에 대해 검토

❒ Development of a method for ex-post identification of falsification in survey data

✍ Christoph Kemper, Natalja Menold, Nina Storfinger and Peter Winker

◦ 이 논문의 주요 목적은 서베이에서의 오류 데이터를 사후적으로 찾아내는 것임

- 일련의 연구를 통해 오류에 민감한 조사표 특성(지표)을 확인했고, 이러한 지표들을 사용하여 위험성 있는 조사원을 확인하기 위해 다변량 통계방식을 적용

◦ 속이는 조사원을 검침하기에 유용한 조사표의 특성(예를 들면 이상한 응답행동, 응답일관성, 항목무응답 등)을 확인

- 첫 번째 연구에서 8명의 오류응답자가 독일 사회일반 서베이에서 뽑힌 하나의 표본에 대해 37개의 거짓 응답 생성

- 두 번째 연구에서 105명의 진짜 응답자에 대해 참가자의 11개 사회인구학적 특성(성, 연령, 교육, 종교 등)에 관한 정보를 제공함으로써 거짓 정보를 생성

- 진짜 응답자는 기존 서베이 데이터(ALLBUS and International Survey Program)에서 확률적으로 뽑음. 진짜와 가짜 응답의 차이가 나는 여러 지표 발견

◦ 발견된 특정의 지표를 통하여 조사원을 옳은 그룹과 속일 가능성이 있는 그룹으로 구분 분석

- 분석 결과, 잘못된 조사원의 비중은 클러스터 분석에 사용된 지표에 의존한다는 것으로 나타남

- 어떤 단일 지표의 경우 약 80%의 속임을 확인하게 해주었음

❒ Social desirability bias in survey measures of illicit work

✍ Mark Trappmann, Ivar Krumpal, Antje Kirchner and Ben Jann

◦ 전화 서베이에서의 민감한 주제에 대한 응답결과를 어떻게 평가할 것인가가 중요한 과제

- 본 연구는 불법노동 서베이에서의 사회적 바람직성 편의를 측정하고자 함

- 어떤 확률화 과정도 필요치 않고 CATI 조사에 쉽게 구현될 수 있는 방식 개발

◦ 이 방식은 실제 행정자료를 통해 알고 있는 값과 비교함으로써 민감 질문에 대한 응답편향을 추정하는 것임

- 우선 2부류의 특정 인구 응답자 (피고용자 집단과 기본적 소득지원을 받을 수 있는 집단) 표본을 행정자료에서 구함

- 새롭게 개발된 방식과 표준적인 직접질문 방식중 어느것이 2 집단에 대해 불법노동의 맥락에서 더 높은 정도의 사회적 바람직성을 나타내는가를 평가

주제6 Access Panels

❒ Mode experiments in the PPSM panel

✍ Edith de Leeuw and Joop Hox:

◦ 혼합방식 서베이가 대안적인 조사방식으로 각광받고 있으나 아직 조사방식에 따른 결과 차이에 대한 연구가 충분치 않음

- 다른 방식에 의해 수집된 데이터는 각각의 조사방식이 다른 조사 과정을 거치고 또한 방식마다 다른 유형의 질문들이 이루어지기 때문에 다른 결과 예상됨

◦ 이 논문의 목적은 조사방식에 의해 거의 영향받지 않을 질문 양식을 선택하여 전화서베이와 웹서베이 결과를 효과적으로 비교하는 것임

- 일련의 실험을 통해 2가지 서베이 방식으로 다양한 형태의 질문을 실험

- 이러한 실험을 통해 데이터 수집 방식간 혹은 그 방식 내부에서의 다른 질문형태간의 효과를 비교

- 데이터는 확률적으로 웹과 전화 패널로 선정된 PPSM 패널 응답자에게 질문하여 수집

❒ The German access panel and the impact of response propensities

✍ Boyko Amarov, Ulrich Rendtel, Tobias Enderle, Ralf Münnich and Stefan Zins

◦ 가구와 개인 서베이에서 심각한 무응답 편향을 발생시키는 단위무응답 문제를 해결하기 위한 하나의 방법으로 자발적 패널(Access panel)을 사용

◦ 이 경우 보다 높은 수준의 응답률이 예상되지만 표본지원 과정의 자기선택 문제는 추정치의 조정을 필요로 함

- 해당 변수를 인구통계에 벤치마킹하거나 추가적으로 응답 성향점수를 이용하는 방식을 통해 조정이 이루어짐

◦ 독일 마이크로센서스는 응답자의 자발적 참여로 이루어지는 독일 국가통계인 DSP(Dauerstichprobe)의 표본 틀 기능

- DSP로부터 도출된 서베이 예를 들면 유럽연합 소득 및 삶의 조건 통계(EU-SILC)의 추정시 단계별 표본 선택과정이 고려되어야 함

◦ 이 논문은 앞선 브루크 등의 Accuracy of Estimates in Access Panel based Surveys(2009)와 아마로프 등의 The Access Panel of German Official Statistics. An Analysis of Recruitment, Panel Attrition and Survey Nonresponse(2009) 연구결과에 기초

- 선택단계의 가중치를 계산하고, 수준과 시간에 따른 변화를 측정하는 데 있어 가중치의 영향을 분석

❒ Building an online immigrant panel: response and representativity

✍ Annette Scherpenzeel

◦ LISS 패널(Longitudinal INternet Studies for the Social Science)은 가구 확률표본에 기초한 웹 패널인데, 2010년 10월부터 LISS 패널은 외국인을 표본에 포함

◦ 이 논문의 목적은 문화적응 과정을 연구하고 외국인의 확률표본 패널을 구축하고 유지하기 위한 방법론에 대한 통찰력을 얻고자 하는 데 있음

◦ LISS 외국인 패널 작성자들은 2010년 이후 웹조사표를 어떻게 설계할지에 대해 고민해 왔음

◦ 네덜란드 인구등록자료에서 뽑은 표본은 인종별로 구분되어 이를 대상으로 자원자 모집은 2010년 5월과 12월에 이루어짐

- 자원자 모집은 LISS 패널의 초기 지원자중 외국인 집단에 대한 경험을 기초로 하여 비서구 인구집단에 대해 수행됨

- 또한 각 인종집단의 응답률은 자원과정 동안 지속적으로 점검되었고, 특정집단에 대해 필요한 경우 수정됨

- 이러한 응답에 따른 조정과정에도 불구하고 3개의 비서구 인종집단이 응답률이 많이 떨어졌고, 패널에 대해 대표성을 갖지 못하게 되어 이 특정집단에 추가 표본을 확보

◦ 수집된 데이터는 LISS 패널데이터의 경우와 같이 연구자들이 이용할 수 있음

- 또한 연구자들은 이 외국인 패널에서 모아진 데이터를 활용하는 제안서를 제출하여 연구지원을 받을 수 있음

❒ Long-term effects in the PPSM-Access Panel

✍ Simone Bartsch, Uwe Engel and Helen Vehre

◦ 이 논문에서는 PPSM 구축의 장기효과를 분석

- PPSM 패널은 2009년부터 확률기반으로 표본이 모집된 일종의 자발적 패널임
- 표본모집을 위해 대상이 되는 응답자를 설득하고 거절 전환한 사람들을 평가하기 위한 특별한 노력을 기울임
- 연구자들이 주장하듯이 응답률 뿐 아니라 표본구성의 관점에서 표본으로 참여하도록 하게 하는 노력과 거절전환의 장기효과를 평가하는 것이 중요

◦ 이 논문은 다음과 같은 2가지 질문을 중요하게 제기

- 최초 설득전략의 장기효과가 무엇인가?
- 설득되어야만 하는 사람들이 다른 패널 구성원들과 다르고 그로 인해 표본구성의 변화에 기여하는가?
- 이 분석은 사회인구학적 변수에 대해 초점을 맞추는 동시에 다른 서베이 관련 항목(사회태도, 만족도)도 다루고 있음

주제7 Sampling

❒ Impact of stratification in choice-based sample studies with double sampling

✍ Sigrid Behr und Walter Schill

◦ 이중추출 혹은 2단계 방식은 누락 혹은 부정확한 변수정보로 수행하는 연구를 위해 개발됨

- 이 연구는 모수 추정을 위해 1단계에 수집된 정보를 활용하는 다양한 층화 전략의 영향을 파악하는 것임

◦ 3개의 층화전략이 선정됨

- 첫 번째 전략은 변수들을 단순 화하는 것. 두 번째는 정보가 없는 층을 없애고 추가적이고 보다 정확한 변수 정보를 포함시키는 것. 세 번째는 1단계의 우도함수가 최대값이 되도록 하는 방식에 근거하여 층화하는 것

◦ 3개의 층화전략을 2개의 질병실험에 적용

- 첫째 실험은 폐암환자의 직업적 위험요인에 관한 경우로서 석면의 누적효과를 평가하기 위해 모든 참여자의 20%에 해당하는 표본에 대해 1단계에서 흡연정보와 석면노출 기간의 정보가 2단계에서 3가지 층화방식을 사용하여 분석됨

- 두 번째의 실험은 구강 치료후 심각한 빈혈의 위험을 검토

◦ 연구 결과 층화에서 포함된 1단계 정보의 양과 정확성이 2단계 분석 결과와 관련이 있는 것으로 나타남

- 가능한 한 관심변수와 연관이 있는 변수의 1단계의 정보를 많이 사용할수록 가장 정확한 모수 추정치를 구할 수 있음

주제8 Linking Survey Data

❒ New items for self-generated identification codes

✍ Tobias Bachteler und Rainer Schnell

◦ 종단 서베이에서 개별응답자의 데이터를 서베이 회차간에 연결할 필요가 있음

- 민감한 주제에 관한 서베이에서 행정 목적으로 사용되는 이름이나 식별수자를 묻는 것은 익명성 유지 원칙에 위배됨
- 이 경우 자기작성 식별코드 방식이 서베이 회차간 응답자를 연계하는 데 종종 사용됨

◦ 자기작성 식별코드는 응답자에게는 알려져 있고, 연구자들은 모르는 요소들로 구성됨

- 예를 들면 자기 이름, 친구 이름의 앞글자, 생일의 일정 부분, 태어난 곳의 앞글자 등이 전형적 예
- 응답자는 각 패널회차 설문의 내용일부로서 같은 코드를 입력하고, 연구자는 각회차의 패널데이터 기록을 자기작성된 식별코드에 기초하여 연계

◦ 그러나 이러한 자기작성 식별코드 방식은 실제 연계되는 비율이 매우 적고, 잘못 연결되는 경우를 발생시키는 문제 존재

- 그동안 자기작성 식별코드로 사용될 수 있는 완전히 새로운 방법을 체계적으로 연구해 옴
- 실험을 통해 새롭게 작성된 코드와 기존의 코드를 비교하였고, 항목과 코드의 타당성은 적용가능성, 엔트로피, 안정성 지표를 통해 평가함

❒ Linking survey data with administrative employment data

✍ Manfred Antoni

◦ 이 연구는 독일 서베이인 ALWA와 연방고용국의 행정자료의 연계결과를 비교

- ALAW 서베이는 1956년과 1988년에 태어난 10,177 표본의 기억기반 인터뷰를 포함. 종단 정보는 거주지, 교육, 고용, 결혼관계, 자녀, 부모 사망 등에 대해 월마다 조사

- 행정자료는 고용이력에 관한 매일의 정보와 이전지불, 임금에 관한 정보가 포함되어 있고, 사회보장기여와 연관되기 때문에 아주 정확히 기록됨

- 자료 연계 동의는 응답자 92%(9,531) 수준

- 고용상황은 경제부문, 피고용자의 자격, 나이, 임금분포, 직업이동 등에 관한 기업정보로 보완

◦ 연구의 주요 내용은 다음 2가지

- 첫째로 데이터 연계과정의 성공은 양 데이터에서 연계되는 것을 발견하기 위한 전략(정확연계 대 확률연계)에 어떻게 의존하고 있는가를 보여줌

- 둘째로 데이터 연계에 대한 응답자의 동의 실제 연계과정의 성공이 응답자와 조사원의 특성 및 인터뷰 상황에 의해 어떻게 달라지는지를 서베이 자료와 패러데이터에 기초한 다변량 프로빗 회귀모델을 통해 분석

Ⅲ 엥겔 교수와의 업무 협의

(1) 회의 개요

◦ 시간 : 2011.`11.15 2:00~4:00

◦ 장소 : EMPAS 연구소

◦ PPSM 프로젝트 추진 현황과 서베이 방법론 연구 동향

(2) 회의 내용

❒ 소개

◦ 한국 통계청은 통계품질을 강화하기 위한 방법론 연구 확대를 지속적으로 도모해 옴

◦ PPSM과 같은 유형의 프로젝트를 중장기적으로 추진하려고 그동안의 경험에 대해 듣기 위해 방문

◦ 궁극적으로 이를 통해 서베이 방법론을 연구하는 연구자 집단(pool)을 유지하고 이들을 연구인력으로 활용해 나가려는 목적

- 서베이 방법론에 대한 청내외 수요가 증대되는 추세임에 비해 내부 인력 부족 및 확대 가능성 취약

◦ 엥겔교수는 이러한 취지와 생각이 바람직하다는 의견을 표명. 이하에서 엥겔교수와 토의한 내용을 정리

❒ PPSM 추진의 계기와 내용

◦ 2007년에 엥겔 교수가 주축이 되어 몇몇 교수들과 서베이 방법론

연구를 중장기적으로 추진하기 위한 펀드를 마련하기 위한 프로젝트 구상

◦ 미국 등에 비해 서베이 방법론 연구 및 적용이 뒤떨어져 있다는 인식하에 독일 국내 및 영국 네덜란드를 포괄하는 서베이 방법론 네트워크 구축

- 2008년 5월~2012년 4월 동안 제공되는 연구기금(672만 유로)

◦ 관련 주제는 이전에 여러 교수들이 모여 표본, 추정, 무응답 등을 중심으로 집중해서 연구해야 것을 논의하여 선정

◦ 이 결과를 바탕으로 관련 기금을 운영하는 연구선정 위원회에 연구 프로젝트를 제안 과정을 거쳐 단계적으로 선정

- 교수들의 주제 제안을 받아서 경쟁 과정을 거쳐 선정하는 것이 특징

- 14개 연구집단이 참여하여 13개 연구주제 추진

◦ 2013년 이후 현행 프로젝트가 마무리되면 다른 기관을 통해 펀드를 확보할 계획

❒ 국가통계와의 관련

◦ 연구주제가 민간에서 활용되는 서베이 방법론이 중심을 이루고 있고, 국가통계에 직접 적용되는 연구 분야는 적음

◦ 노동통계에서 행정자료와의 연계를 제외한 대부분의 주제는 독일 통계청과 직접적인 관계가 적음

◦ 그러나 PPSM에서 다루어지는 무응답 문제, 표본설계, 자발적 패널 등의 연구주제는 향후 국가통계 분야의 중요한 서베이 방법론으로 개발 적용될 것이라는 데 공감

❒ 연구자원

◦ 미국에서는 통계학 기반의 서베이 방법론 연구자들이 많은 데 비해 유럽 특히 PPSM을 수행하고 있는 연구자들은 통계학보다는 사회학 계통 혹은 사회과학 분야가 다수를 차지

◦ 독일의 경우 사회학자가 중심이 되어 서베이 방법론 연구를 주로 수행해 왔으며, 이론 기반의 순수 통계학자의 참여도가 낮은 편이나 상호간의 학제적인 추진이 바람직하다고 판단하여 서베이 방법론 이론에 강한 통계학자들의 참여를 유도하고 있음

◦ 그동안 연구 주제를 선정하고 연구자문위원회에서의 경쟁 심사 과정을 거치는 데도 사회학자 등의 시야에서 서베이 방법론의 문제와 연구방향을 제시하는 것이 더욱 현실적이었음

❒ 국제 연구협력 및 인력 양성

◦ 그동안 주로 유럽지역을 중심으로 공동연구가 진행되었고, 아시아지역 통계청이나 연구자(학자)들과의 교류 경험이 없었다고 함

◦ 적절한 프로젝트 형태로 원하는 주제에 대해 원하는 연구자를 요청할 경우 공동연구 등의 협력을 추진할 의사가 있다고 함

- 특히 젊은 유망한 서베이 방법론 연구자 교류 고려

◦ 내년에 전문 서베이 방법론 주제로 한 단기교육 프로그램이 연구자 양성기관인 GESIS(라이브니치 사회과학 연구소)에서 대학원 학생들을 대상으로 운영될 계획인데, 한국 통계청의 참여 요청이 있을 경우 연결시켜 줄 것임

◦ 독일, 네덜란드, 영국의 경우 서베이 방법론을 연구하려는 학생들이 이 분야의 중요성에 대한 인식 증대로 인해 증가하는 추세

IV 출장 소감

❒ 서베이 방법론 연구 최근 동향 파악

◦ 본 회의에서도 금년도 유럽 지역의 서베이 방법론 모임(ITACOSM, ESRA)이나 미국의 JSM과 유사한 주제들(무응답, 자료통합, 혼합조사, 자원자 패널 등)을 다루고 있다는 점에서 최근 서베이 방법론 연구의 초점이 어디에 있는가를 확인

- 이 주제들은 통계개발원의 향후 중점연구과제 분야이기도 함

◦ 동시에 출장자들은 본 회의에서 다음과 같은 몇가지 특징을 파악

- 통계적 실험의 적용 : 대부분의 연구가 조사를 보완하는 실험을 통해 방법론적 이슈를 해결하려는 시도를 하고 있음

- 유럽에서는 추정보다는 조사단계의 정확성을 확보하기 위한 연구가 활발히 이루어지고 있는 것으로 판단됨
  · 통계학자보다는 사회학이나 경제학자가 조사의 정확성을 주로 다루려는 경향과 연관이 있는 것으로 보임

- 범유럽 연구네트워크의 구축 : 주요 방법론 이슈에 대해 유럽 주요 연구기관들이 서로 연계하여 공동으로 연구하는 양상
  · PPSM은 독일, 네덜란드, 영국을 중심으로 한 공동연구 네트워크의 성격을 가지며, 유사한 조직으로 방법론 전문 연구기관인 GESIS의 관련 분야 활동도 눈에 띔

- 방법론 연구 인력의 확대 추세 : 통계조사의 환경 악화에 따른 방법론 개발의 중요성에 대한 인식이 증대되면서 유럽지역 서베이 방법론 연구 인력이 증가하는 추세이고, 이를 위한 교육기관(GESIS)의 활동도 활발

❒ 통계청의 서베이 방법론 연구 발전을 위한 시사점

◦ 통계조사의 정확성을 제고하기 위해 조사단계에서의 편향(bias) 발생 가능성에 대한 실험 연구 수행이나 파라데이터 확보 필요

- 소규모의 실험을 통해 조사단계에서의 오류 발생 가능성을 탐지하기 위한 연구나 파라데이터의 확보를 통해 조사의 품질을 제고하기 위한 연구 수행 필요

◦ 통계학 전공자와 사회학, 경제학 전공자의 공동연구에 의한 서베이 방법론 연구 보완 필요

- 통계학 전공자는 주로 조사 이후의 추정쪽에 연구의 중심이 있고, 사회학·경제학 전공자는 조사표나 조사내용 쪽에 연구의 중심이 있어 종합적 서베이 방법론 개발에서 한계가 존재

- 표본설계, 조사단계에서 발생할 수 있는 편향, 가중치 적용 혹은 조사방식에 따른 차이 등에 대해 학제적 접근이 필요

◦ 국가통계의 영역 확대 및 학계 연구성과의 활용

- 학계의 연구는 주로 소규모 통계를 대상으로 하거나 자발적 패널 등 국가통계에 당장 적용하기 어려운 분야가 중심

- 그러나 점차 국가통계의 영역에서도 기존 조사방법의 한계, 새로운 조사기법의 대두, 다양한 통계수요의 요구 등의 이슈가 더욱 부각되면서 학계의 기존 연구성과의 수용이 필요할 전망

◦ 서베이 방법론 연구의 내부역량 강화와 외부 연구역량 활용을 위한 투자 확대와 네트워크 구축

- 직원 서베이 방법론의 해외연수에 의한 내부 연구역량 제고

- 해외 서베이 방법론 연구자 초청 및 공동연구