

2017년

제61차 ISI 세계통계대회 출장결과

The 61th World Statistics Congress of Int'l Statistical Institute

2017. 8.



기 획 조 정 관
국 제 협 력 담 당 관

● ● ● **차 례** ● ● ●

I. 출장 개요 3

II. 주요 활동 4

III. 우리청 참가논문 발표내용 요약 6

IV. 전문세션 참가내용 요약 8

V. 건의사항 15

【참고1】주요 활동 사진 16

【참고2】전문세션 참가 결과 자료 17

I 출장 개요

- 출 장 목 적 : 제61차 ISI 세계통계대회 참석
- 기 간 / 지 역 : '17. 7. 16~7.21 / 모로코 마라케시
- 참 석 자 : 안형준 경제통계국장(단장)외 4명

구분	성명	소속(직급)	업무수행내역
단장	안형준	경제통계국장	· 제61차 ISI참가 대표단장 · 대표단 대외협력활동 업무 총괄
논문 발표	임성주	국제협력담당관실 (5급)	· ISI 본회의 준비 및 참가, 단장 보좌 · 발표주제 : “소비자물가지수 작성산식 검토” · 타 세션 참가 및 자료 수집 및 결과보고
	박은영	물가동향과 (5급)	· 발표주제 : “개인맞춤형, 나의 물가계산 시각화 콘텐츠서비스” · 타 세션 참가 및 자료 수집
	박윤영	동남청 지역통계과 (5급)	· 기고논문발표(울산베이비부머 통계개발과 지역정책 수립 모니터링) · 타세션 참가 및 자료 수집
	황수린	빅데이터통계과 (6급)	· 발표주제 : “공공 빅데이터를 이용한 공동주택 초등학교학생 유발률 계수 개선” · 타 세션 참가 및 자료 수집

○ 출장일정

날짜	출발지	도착지	비행시간	업무수행 내용
7.15.(토)	인천 (11:00)	샤를드골 (16:00)	12h 00m	· 출국 (항공편 : KE0903)
	오를리 (21:15)	마라케시 (23:20)	3h 05m	(항공편 : AT0645)
7.16.(일)	모로코 마라케시(Marrakech)			· 등록(09:00~16:30) · 환영식(16:30~19:00)
7.17.(월)				· 세션 참가
7.18.(화)				· 세션 참가 · IAOS 총회참가
7.19.(수)				· 세션 참가 · IASS 총회참가
7.20.(목)				· 세션 참가 · ISI 총회참가(17:15~19:15)
7.21.(금)				· 세션 참가 · 폐회참가 · 환송식 참가(20:30~23:30)
7.22.(토)~ 7.23.(일)	마라케시 (10:05)	오를리 (14:10)	3h 05m	· 귀국 (항공편 : AT0640)
	샤를드골 (21:00)	인천 (15:00)	11h 00m	(항공편 : KE0902)

1. 우리 청 참가 논문 발표 및 질의 응답 실시

[논문발표 일정]

구분	발표자	발표일	소속(직급)	발표주제
기고	임성주	7.21	국제협력담당관실(5급)	소비자물가지수 작성산식 검토
기고	박은영	7.17	물가동향과(5급)	개인맞춤형, 나의 물가계산 시각화 콘텐츠서비스
기고	박윤영	7.20	동남청 지역통계과(5급)	울산베이버버머 통계개발과 지역정책수립 모니터링
기고	황수린	7.20	빅데이터통계과(6급)	공공 빅데이터를 이용한 공동주택 초등학교 유발률 계수 개선

2. 주요행사 및 전문세션 참가를 통한 정보수집

- (주요행사) ISI 현재 활동 점검 및 방향설정을 위한 종합 행사 등
- 총회*, 환영식, 환송식, 시상식, 폐회식, ISI 회장 특강, IAOS 및 IASS 주관 행사 참가

* 총회(7. 20일 개최) 안건으로 상정된 활동보고서, 재무보고서, ISI 산하 협회 활동보고서, 세계통계대회 개최 및 유치(2017 Marrakech, 2019 Kuala Lumpur, 2021 네덜란드)에 대해 심의 및 승인 동의

- (전문세션) 국내외적으로 부각되고 있는 공식통계개선, 통계역량 강화 등 관심 세션 참가하여 정보 수집

* (IPS100)통계적 비즈니스등록부를 위한 국제가이드라인의 개발과 적용
(IPS060) 주목을 받고있는 데이터 사용자-통계 활용
(IPS104) 공연하는 통계학자가 된다는 것
(IPS021) 국가중앙은행과 통계청 간의 통계협력 강화
(IPS078) 설문조사 무응답 처리하는 혁신적인 접근법
(IPS053) 공식통계의 생산을 별도의 정부부서에서 할 필요가 있는지 혹은 정부부서내에 남아 있어야 하는가?

- 아울러, Big data 관련 주요 세션에 참가하여 최신 동향을 파악

* (IPS 136) 빅데이터와 큰 기회들
IPS 012) 최근 발전되고 있는 빅데이터 공동모델
IPS 024) 빅데이터시대의 공식통계

3. 국제 교류 활동

[말레이시아통계청]

- 말레이시아측은 2019년 개최시 대회장 소개, 유치계획 등 사전 설명회를 실시
 - * 2015년 ISI 총회에서 유치지 승인을 받고, 2017년 동 행사에서 공식 초청(2019 ISI PPT 및 동영상) 실시

[일본통계센터 전시관 방문]

- 일본측은 2018. 7월 일본에서 개최 예정인 제10차 통계교육 컨퍼런스*에 대하여 설명
 - * ISI 산하 분과위원회의 하나인 통계교육연구회(IASE) 주관으로 일본통계센터 및 통계국 주관으로 개최

[기타]

- 예정되었던 아부다비와의 양자협의를 아부다비측 사정으로 미개최

1. 소비자물가지수 작성산식 검토

(7.21. 14:00~14:50 기고논문 조사통계세션, 발표자 국제협력담당관실 임성주)

- (내용) 현재 통계청은 CPI를 라스파이레스산식을 사용하여 기준년 개편시(중간년도 가중치 개편 포함)마다 23개월정도의 변동률의 보정이 이루어지고 있어 이용자(연금, 계약 등)들의 불편함이 상존
 - 현행 라스파이레스 산식과 로우산식, 해외사례*를 비교하고, 산식 변경시의 검토사항(CPI 품목수와 가계동향항목간의 일치 등)을 설명
 - * 미국, 유럽 등 선진국들은 로우산식을 적용, 개편시에도 변동률 미수정
- (결론) 다양한 물가연동 파급효과를 고려할 때 지수의 안정성에 대한 요청이 매우 강조되는 상황하에서, 향후 지수산식변경(Laspeyres→Lowe)을 통해 소급 없이 신지수를 도출할 필요성이 있음

2. 맞춤형 개인물가지수 시각화 콘텐츠 서비스

(7.21. 16:00~17:40 기고논문 Method and Theorys 세션, 발표자 물가동향과 박은영)

- (내용) 한국의 경우 최근(2014년-2016년) 2%미만의 소비자물가상승률을 보이는 반면 일반국민이 체감하는 물가는 2%를 상회하여 간극이 커지고 있어 공식통계에 대한 불신을 해소하기 위해 각 개인의 소비패턴을 반영한 나의 물가지수 계산 콘텐츠를 2016년에 개발하여 대국민 서비스하고 있음을 설명
 - ‘나의 물가 체험하기’는 소비자물가조사 전체 460개 품목 중에서 개인이 소비하는 품목만 선택하여 계산하는 ‘기본형’과 지출액의 비중도 반영하여 계산하는 ‘도전형’으로 고안됨
- (결론)각 개인이 소비하는 품목과 지출액을 반영함으로써 나만의 맞춤형 물가지수를 공식통계(CPI)와 비교함으로써 차이를 실감할 수 있도록 국민들의 이해를 제고함

3. 울산베이비부머통계 개발과 지역정책 모니터링

(7.20. 14:00~15:40 기고논문 공식통계세션, 동남청 지역통계과 박운영)

- (내용) 베이비부머의 은퇴가 본격 실시됨에 따라 울산시의 경우 지속 가능한 도시 경쟁력 제고를 위한 정책개발을 위해 동남지방통계청과 함께 베이비부머 통계를 개발
 - 울산지역 베이비 부머의 인구특성, 건강 및 보건, 재정상태, 의식 등 약 70여종의 통계표로 이루어진 주요결과와 아울러 통계가 다양한 행정자료 및 자료처리 과정을 통해 정제되어 만들어 지는 통계제작과정을 소개
- (결론) 만들어진 통계가 합동 설명회를 거쳐 지역정책부서의 통계 정책 연계 Matrix 과정을 통해 정책이 도출되는 과정과 만들어진 통계기반 정책형성의 내용을 지역단위에서 보여줌으로써, 지역통계가 정책을 선도하는 Best Practice의 한 사례로 주목을 받음

4. 등록센서스 및 인구주택총조사 자료를 이용한 초등학교 학생유발률 계수 개선

(7.20. 14:00~15:40 기고논문 Applications 세션, 발표자 빅데이터통계과 황수린)

- (내용) 최근 저출산으로 인한 학령인구 감소와 도심내 지역별 학생수 양극화로 인해 적정규모의 학교설립과 학생재배치 문제가 중요한 이슈로 떠오르고 있음
 - 2010 및 2015 인구주택총조사를 활용, 보다 정확한 학생수 예측을 위해 기존 광역시 전체 계수만 사용했던 학생유발률 계수를 신·구도심별/읍면동별/연면적별로 세분화하여 분석사례 설명

* 초등학교 학생유발률 = $\frac{\text{초등학교 학생수}}{\text{해당 지역의 주택수}}$ ** 예상 학생수 = 주택수 * 학생유발률

- (결론) 학생유발률 계수의 개선을 통해 지역적 격차해소 및 적정규모의 학생수를 예측하는데 도움이 될 것으로 기대

[공식통계개선 및 통계역량 강화 세션]

1. (IPS 100) 통계적 비즈니스등록부를 위한 국제가이드라인의 개발과 적용

○ 세션요지

- 동 세션에서는 유럽통계처, UNECE, AFDB, ADB 에서 추진하고 있는 기업등록부 매뉴얼에 대한 4개 기관별 추진내용을 설명하고 토의가 진행되었음. 유럽통계처 발표자는 유럽연합 기업등록부의 작성배경, 주요내용, 개선계획에 대한 종합적인 내용을 설명하였음
- 두 번째 발표자인 UNECE 발표자의 경우 유럽통계청에서 활용하고 있는 통계기업등록부에 대한 내용을 설명
- 세 번째인 아프리카 개발은행 발표자의 경우 아프리카 지역 현실에 맞춘 기업등록부 지침서 제작경위와 실제활용국가에 대해 사례발표를 실시
- 마지막 발표자인 ADB 직원의 경우 기업등록부의 기본규정과 실제 시험적으로 작성되고 있는 나라들에 대한 소개를 실시.

○ 세션시사점

- 기업등록부의 가이드라인, 이행, 지속적 개선은 국제적인 공동 관심사 안으로 한국형 통계기업등록부 역시 이러한 국제기구들의 움직임에 상세히 파악하고 국제비교가능한 통계기업등록부가 될 수 있도록 하여야 할 것임
- 기업등록부에 관한 비스바덴 국제회의 유치 혹은 국제기구 및 선진국 전문가 초청 세미나 개최 등을 통해 BR의 활용가능성을 높이고, 한국형 기업등록부국제표준을 선도해 나갈 수 있는 기회의 장을 마련해 보는 것도 바람직해 보임

2. (IPS 060) 주목을 받고있는 데이터 사용자-통계활용

○ 세션요지

- 통계의 중심을 이용자관점으로 옮겨놓아야 한다는 점에 주안을 둔 세션으

로 아일랜드 통계청, 핀란드통계청, 유럽통계처에서 추진하고 있는 이용자 중심 서비스 전략에 대한 다양한 논의가 있었음

- 핀란드 통계청은 고객이 조직전략속에 보여야 하고, 통계개발과정에 고객과 이해관계자가 참여되어야 하고, 전체적인 고객의 평가 및 환류시스템이 조직운영관리에 포함되어야 함을 강조함.
- 유럽통계처의 경우에도 디지털의사소통 프로젝트(DIGICOM)의 일환으로 유럽통계시스템(ESS)이 고객중심으로 설정되어야 함을 피력하였음. 디지콤 프로젝트는 이용자 지식 개선, 이용자와의 관계강화, 신규이용자에게 다가가기, 이용자 관점의 프로젝트 진행으로 구성되어 있음

○ 세션시사점

- 유럽통계처 및 유럽소재 통계청의 경우 이용자 중심의 통계서비스 전략 수립에 분주하고 있음을 볼 수 있음. 고객들의 목소리를 그대로 반영하여 서비스를 개선하고 이를 통해 만족도를 제고시키는 모습이 상당히 인상적임
- 유럽통계처의 경우에도 유럽통계올림픽 등 이벤트 행사를 통해 이용자의 체험 기회를 확대하고자 하는 아이디어 등을 적극 발굴하는 모습을 벤치마킹할 필요성이 있음

3. (IPS 104) 공연하는 통계학자가 된다는 것

○ 세션요지

- 발표자인 David Spiegelhalter는 캠브리지 교수로서 다양한 매체와 딱딱한 통계를 쉽고 재미있는 콘텐츠로 만들어서 전달하는 다양한 시도들을 보여주고 있음. 자칫 딱딱해서 통계를 싫어하는 이용자들에게 음악, 미술 등 문화콘텐츠처럼 통계도 대중화 되었으면 하는 바람을 강연 내내 표명하였음
- David교수는 방송사에서 수영장을 배경으로 제작한 서바이벌 게임에 직접 출연하여 실패 및 성공을 통계로 수치화하여 시청자에게 몸으로 소통하는 통계 전달을 시도함. 아울러, 눈사람과 자신이 하늘을 나는 CG를 만들어 왜 미디어와 통계인이 만나야 하는지 다양한 설명을 시도. 우선 새로운 스토리에 대한 평가시도, 교육, 통계과학 로비, 연구와의 소통, 공공의 관심을 유도 하는 것 등이 그에 대한 이유라고 설명.

○ 세션시사점

- 통계작성자 중심의 통계생산 패러다임에서 벗어나 이용자를 중심으로 통계가 만들어져 용이하게 사용되는 환경 인프라 조성을 강조되고 있음. 통계서비스를 중심으로 통계의 축이 변화되고 있음을 감지
- 이용자의 층을 다양하고 세밀하게 분석하고 계층별 촘촘한 이용자 만족 서비스 개발 및 통계이용 극대화를 위한 다양한 전략책 강구의 필요성 실감

4. 국가중앙은행과 통계청 간의 통계협력 강화

○ 세션요지

- 세션은 중앙은행 경제학자와 통계학자뿐만 아니라 중앙은행이 관심을 갖는 통계적 쟁점에 대해 토론하기를 원하는 포럼 기구인 IFC(Irving Fisher Committee on Central Bank)가 구성한 세션임
- 전 세계적으로 거시 경제 통계 자료는 여러 기관에서 제공하여 작성되는데 특히 각국 중앙은행(Central Bank, CB)과 국가 통계청 (NSO)이 생산하는 통계가 가장 많으므로 두 기관 사이에는 표준 및 방법론을 개발하기 위한 상당한 노력, 통계들 간의 일관성을 보장하고, 두 기관 간 협력 증대가 필요

○ 세션시사점

- 이 세션은 각국 중앙은행이 중심이 되어 구성된 세션이므로 중앙은행 입장에서 각국통계청과의 협력과 협조가 중요함을 제시하는 내용으로 구성되었으며 우리나라 한국은행 경제통계국 금융통계부 부장이 의장을 맡음
- 각국 중앙은행의 발표자중 Eurostat의 John Verrinder는 중앙은행과 각국 통계청간의 관계를 'Wife and Husband'.로 표현했는데 이는 '가깝고도 먼 사이' 라는 것을 의미한다고 설명.
- 우리나라의 경우도 이런 가깝고도 먼 관계가 아닌가 싶었으며 그럼에도 불구하고 통계작성의 효율성이라는 측면에서 통계청과 한국은행이 서로 협력할 수밖에 없는 관계임을 인식하고 협력의 필요성을 인정해야 할 것임. 특히 GDP 중심 국민계정을 작성하는 한국은행과 지역국민계정 (GRDP)을 작성하는 통계청이 상생의 관점에서 서로 협력할 필요성이 있다고 보임. 한편 이 세션은 각국 중앙은행 중심의 발표인데 각국 통계청도 함께 참여함으로써 서로간의 협력을 위한 상호의 관점에서의 발표와 토론이 이루어질 필요가 있다고 생각함

5. (IPS078)설문조사 무응답 처리하는 혁신적인 접근법

○ 세션요지

- 가계 조사의 정확한 무응답률을 계산하기 위해 조사과정에서 수집된 실시간 paradata를 이용하는 방법을 제시함. 조사 관리자는 설문 조사 프로세스를 모니터링하고 응답자의 응답률이 전반적으로 낮아지는 경우 필요한 조치를 취할 수 있음. 파라데이터를 통해 수집된 자료의 품질을 평가할 수 있는 여러 리서치들이 있어왔고 세계은행의 'survey solution' 이라는 파라데이터 수집 솔루션을 통해 응답률을 예측하고 목표 응답률을 높이는데 활용될 것이며 조사 과정의 효율성을 높여 응답률을 극대화 할 수 있을 것임. 그러나 현재 이러한 방법은 여전히 기술적인 문제, 법적인 문제 등이 놓여 있지만 관련된 문제를 하나씩 풀어나갈 계획

○ 세션시사점

- 통계조사 과정에서 무응답을 줄이는 노력은 다양한 분야에서 시도되어 왔음. 우리나라도 현장 방문을 통한 조사자료를 행정자료로 대체함으로써 예산절감, 응답부담 감소 및 무응답 감소 등으로 통계의 내적 및 외부적 측면에서 품질과 효율성을 높여왔음.
- 조사과정에 관한 자료로서 파라데이터 수집 및 분석에 관한 논의는 우리청에서도 몇 년 전부터 시도되어 왔고 일부 모바일조사의 경우 별도로 조사하지 않아도 파라데이터가 수집 가능함. 그러나 좀 더 실질적인 활용 사례에 대해서는 아직 연구가 더 필요한 것으로 보임

6. (IPS053)공식통계의 생산을 별도의 정부부서에서 할 필요가 있는지 혹은 정부부서내에 남아 있어야 하는가?

○ 세션요지

- 전 호주 통계청장인 Dennis Trewin이 '독립적 공식통계작성기관이 실질적으로 의미하는 것은 무엇인가?' 라는 제목의 발표에서 통계기관에 있어 가장 기본은 신뢰성임을 강조하였고 통계의 신뢰성이란 '통계는 객관적으로 작성되고 발표되어야 하며 통계의 품질에 확신이 있어야 하고 공식통계의 사용은 자격요건이나 질문 없이 사용되어야 하며, 비밀보호는 통계의 신뢰를 유지하기 위한 필수 요건'임을 강조함

○ 세션시사점

- 발표하는 연사 중 2004년에 우리나라 통계청을 방문했던 당시 호주 통계청장이었던 Dennis Trewin인 것이 무척 인상적이었고 공식통계를 작성하는 기관의 핵심요건은 신뢰성(Trust)임을 다시 한 번 느낄 수 있는 세션이었음.

[빅데이터 세션]

7. (IPS136) 빅데이터와 큰 기회들

○ 세션요지

- (공식 통계작성에서 빅데이터를 활용한 전통적 조사방식 대체) 빅데이터로의 접근 및 분석기술 발전, 데이터의 정확성 요구, 건축 예산 등 빅데이터를 활용하여 기존 표본조사를 대체하기 위한 불가피한 압력이 가중되고 있음
- 본 발표에서는 공식통계를 위해 빅데이터를 사용하는 중요한 도전과제를 논의하고, 그 장점과 한계에 대해 언급하고, 두 번째로, 제한된 인구데이터와 빅데이터를 연계하는 통계모델을 제시
- (빅데이터를 이용한 교통패턴 및 여행 이동시간 예측 : 소지역 추정 접근법) ATIS(Advanced Traveler Information System)는 여행 전(前)과 여행 중 사용자에게 충분한 정보를 제공하여 의사결정을 내릴 수 있게 해 줌, 본 발표에서는 소규모 지역 추정(SAE)의 아이디어를 차용하는 새로운 방법을 제안하고, 소지역 추정치가 자동차용 프로브 프로젝트(VPP:메릴랜드 대학 운영) 빅데이터를 활용하여 실시간 교통예측을 실현할 수 있는지에 대해 설명
- (이탈리아 실업률을 이용한 소지역 모델 기반 추정 : Google 트렌드 데이터) 많은 연구자들은 공식 경제지표를 개선하기 위해 온라인 인터넷 데이터 사용을 제안함. 최근 문헌에 따르면, 구글 트렌드 구인광고 시리즈를 활용한 이탈리아 실업률 관련한 다양한 시계열 모델이 제시됨. 이 논문의 목적은 인터넷 검색 데이터 사용을 이탈리아의 소규모 지역실업률 추정까지 확장하는 것임

○ 세션시사점

- 빅데이터의 공식통계 활용은 대표성 부족 등의 품질문제, 모집단의 부재로 인한 추정치 작성 불가 등의 한계가 있으나, 이를 해결하기 위해 각국에서는 이론적, 제도적 노력이 다양하게 진행되고 있음
- 자료수집 비용 및 응답부담 증가, 시의성 있는 신규통계 수요 증가 등 통계작성 환경의 변화로 인한 새로운 자료수집 방법 필요에 따라, 공식통계작성기관에서는 빅데이터 통계를 공식통계로 활용하기 위해 새로운 통계작성 패러다임으로의 변화를 도모해야함

8. (IPS012) 최근 발전되고 있는 빅데이터 공동모델

○ 세션요지

- (대용량처리 DNA메틸화 데이터분석: 평균과 분산의 동일성 결합 검증) DNA 메틸화는 유전자 코드 변경 없이 유전자 발현을 조절하는 유전적 메커니즘으로, 흡연 등 환경적 요인에 의해 수정될 수 있으며, 유전/환경적 요소 및 암과 같은 복잡한 질병에 있어 중요한 역할을 함
- 최근 몇 연구원들이 DNA 메틸화 지수의 분산을 비교하는 것을 제안하였으며, Ahn과 Wang(2013년)은 질병에 걸린 대상과 건강한 대상 두 그룹 간의 평균과 분산을 함께 비교하기 위한 결합 검증을 제안
본 논문에서 더욱 복잡한 질병의 DNA 메틸화 지수를 살펴보기 위해 새로운 joint test를 제안
- (이산, 연속, 반연속 종단데이터의 결합모델링 응용프로그램) 이산, 연속, 반연속 종단 데이터는 생물학, 환경, 보건 사회연구에서 주로 다루는데, 본 연구에서는 분포가 없는 무작위 효과를 갖는 Tweeie 일반 선형 모델을 사용하여 혼합유형의 종단 데이터를 공동 모델링하는 것을 제안함
- (HIV백신 연구를 위한 결합 모델) 최근에 몇 년간 종단 및 생존 데이터에 대한 결합모델이 주목을 받고 있음. 본 발표에서는 HIV백신 연구를 위한 결합모델을 고려하는데, 다음과 같은 문제가 발생할 수 있음.
(i)종단 데이터가 2진 및 연속 데이터와 같은 다양한 유형의 데이터가 혼합될 수 있음 (ii) 수량의 한계로 인해 종단 데이터를 검열 (iii) 종단 방향 데이터는 오류로 측정할 수 있으며 값 누락 가능 또한 우도

비 추론과 관련된 계산은 매우 까다로움. 발표자료에서는 앞서 언급한 문제를 해결하기 위한 결합 모델과 효율적인 계산방안을 제시함

○ 세션시사점

- 의학, 생물학, 보건 사회분야에서의 빅데이터 분석은 기존의 평균 뿐 아니라 분산까지 비교함으로써, 질병의 예방, 진단, 치료에 효과적인 방안을 모색함. 또한 결합 모델을 통해 기존의 연구를 보완하고, 효율적인 산식을 제시가능함

9. (IPS024)빅데이터시대의 공식통계

○ 세션요지

- (공식통계 작성을 위한 빅데이터 기술채택: 기회, 경험 및 교훈) 공식 통계 작성을 위한 빅데이터의 사용은 최근 몇 년간 국제 수준에서 여러 자료원에 대한 접근, 새로운 방법론개발 및 새로운 도구사용과 같은 측면에 중점을 둠.
그러나 병렬 컴퓨팅 플랫폼 및 NoSQL 데이터베이스와 같은 기술통계, 각국 통계청에는 기술이 빅데이터 혁명의 주요 원인이었음에도 불구하고, 주로 실험활동으로 제한됨. 본 발표에서는 빅데이터 기술을 활용한 통계작성 지원 및 이탈리아 통계청 및 국제수준에서의 경험을 분석하는 다양한 방법을 제시함.
- (신용평가를 예측하기 위한 구글 트렌드 사용 평가)예측에 관한 자료들은 보통 구글 검색 데이터를 통해 이루어지고 있는데, 특정기간 동안 구글에서 검색한 키워드는 웹 사이트 구글 트렌드를 통해 실시간으로 무료로 제공. 또한 Google Correlate도 주목을 받고 있는데, 이는 관심있는 사항에 대해 많은 잠재 예측 변수를 사용할 수 있는 장점이 있음.
· 본 발표에서 주택구매를 위한 프랑스 신용 흐름을 예측하기 위해, 구글 트렌드 및 구글 correlate가 유용한지 알아봄

○ 세션시사점

- 빅데이터 관련 통계조직 내에서 IT 전문가와 통계전문가 관계도 중요한 문제로, 서로 배제하지 않는 환경에서 각자의 역량강화가 중요함

1. 국제협력 무대로 보다 적극적 활용이 필요

- ISI는 2년 주기 국제협력의 장(국제기구, 각국 통계청, 학계, 민간 기관, 협회 등)으로 활용 가능
 - 국제기구 인사로는 유럽통계처, UNECE 통계국, ILO·UN·WB 등 고위직 및 중견간부 등이 참석

2. 국제 통계 동향을 파악할 수 있는 준비가 필요

- 약 300여세션의 논문이 발표되는 만큼, 국제 통계 동향을 파악하기 위한 기회로 활용하는 것이 바람직
 - 전문적인 논문이 다수인 만큼 차기부터는 통계개발원을 중심으로 청내논문심사 등 보다 철저히 준비해서 참석할 것을 제안

참고 1

주요 활동 사진

[우리청 참가논문발표]



[말레이시아 대표단, 일본통계센터 등 면담현장]



[주요행사현장]

ISI 등록현장



공식통계연구회(IAOS) 총회



각종 전시부스



폐회식



[(IPS 100) 통계적 비즈니스등록부를 위한 국제가이드라인의 개발과 적용]**□ 세션개요**

- 일시/장소 : 2017. 07. 17. 10:30~12:30 / A.1.16
- 세션구성
 - Organizer/Chair : Michael Colledge
 - Presenter ① : Amerigo Liotti
"Experiences in application of the European Statistical System Business Registers Recommendations Manual"
 - Presenter ② : Abstract of "development and application of the SBR guidelines of the conference of European statisticians"
 - Presenter ③ : Development and application of the African development bank statistical business register guidelines
 - Presenter ④ : Establishing the statistical business registers in Asia
 - Discussant : Michael Colledge
- 참석인원 : 약 70명

□ 세션요지

- 동 세션에서는 유럽통계처, UNECE, AFDB, ADB 에서 추진하고 있는 기업등록부 매뉴얼에 대한 4개 기관별 추진내용을 설명하고 토의가 진행되었음. 유럽통계처 발표자는 유럽연합 기업등록부의 작성배경, 주요내용, 개선계획에 대한 종합적인 내용을 설명하였음. 두 번째 발표자인 UNECE 발표자의 경우 유럽통계청에서 활용하고 있는 통계기업등록부에 대한 내용을 설명하였고, 세 번째인 아프리카 개발은행 발표자의 경우 아프리카 지역 현실에 맞춘 기업등록부 지침서 제작경위와 실제활용국가에 대해 사례발표를 실시. 마지막 발표자인 ADB 직원의 경우 기업등록부의 기본규정과 실제 시험적으로 작성되고 있는 나라들에 대한 소개를 실시.
- 유럽통계처 발표내용을 살펴보면 유럽통계체계(ESS)내에서 기업등록부 권고안 탄생에 대한 내용을 우선 설명. 통계적 기업등록부(SBR)의 설계와 관리는 법령 2186/9301이 채택된 1993년 유럽의 모든 통계청의 의무규정임을 설명함. 기업등록부 권고초안은 2003년에 공표되었고, 현재 버전은 2010년에 만들어진 신 법령 177/2008이 발효된 직후에 제작됨. 특징으

로는, 기업집단수준의 변수들과 단일시장 및 경제의 세계화 진전문제 포괄, 유럽기업 등록부의 제작과 지분관계에 대한 정보관리에 대한 가이드라인 제시, 법령으로 만들어진 유일한 형태의 권고안임. 아울러 법정단위, 기업체, 지역단위, 기업집단이 EU법령에 정의된 방식으로 규정되도록 주안을 둠.



- 권고안의 구조를 보면 기본편과 생멸 단위편으로 구성되어 있고, 기본편에서는 목적, 단위, 내용과 접근(1장~10장)을 다룸. 아울러 권고안의 범위, 기업등록부의 조화, 통계목적의 기업등록부 활용과 목적, 등록부의 유지, 등록의 내용, 등록의 범주, 법정단위와 기업체, 기업등록부에서의 지역단위, 등록부 접근, 품질정책 등도 포함되어 있음. 생멸단위는 변동 및 연속성(11~16, 21~22장)편에 수록되어 있음. 변동을 다루는 일반적인 접근법, 일반적인 생멸사건에 대한 종합 이해, 기업체에 주안을 둔 생멸변동, 기업체에 대한 연속성 규칙, 지역단위에 주안을 둔 생멸변동, 지역단위의 연속성 규칙, 기업집단, 기업집단에 대한 생멸변화와 연속성 규정이 수록되어 있음.
- 권고안에 대한 평가를 위해 2016년 말 유럽통계청은 현재 BR 매뉴얼에 대한 의견을 청취하기 위해 회원국 통계청대상으로 설명회를 실시하였고, 15개국이 응답하였음. 질문은 BR매뉴얼이 주요 목적과 연계되는지 여부, 활용성 여부, 주요개정필요성 등이었음. 규정 이행의 지침서 역할, 필수정보 수록, 내용의 독립성 등이 좋은 평가를 받는 한편 모범사례 수록, 기업집단 소개 등은 다소라는 평가를 받음. 이러한 평가에 기초하여 향후 개정안은 미래법률체계, 품질을 고려하고 실제적용사례를 포함 및 기업집단과 관련된 명확한 지침 수록의 필요성이 요구됨.



- 이어서 발표된 내용은 UNECE의 기업통계등록부(SBR)의 가이드라인에 배경, 가이드라인 개발 목표와 접근방식, 가이드라인 구조, 가이드라인 적용, 향후 계획 등을 설명하였음. 배경으로는 국제적으로 비교가능한 경제통계, 더 통합적이고 상세하고, 일련성이 높은 지침에 대한 필요성 증대, 국제적 경제관계와 기업집단의 중요성, 통합된 경제통계와 통계생산 및 서비스의 현대화에 대한 SBR의 적정성, 기업통계 표본들을 제공하는 전통적인 역할에 추가한 SBR의 새롭고 힘 있는 역할, 2012년에 SBR을 유지하고 만들기 위한 T/F 창설 등을 배경으로 마련된 시안은 2012년, 2014년 비스바덴 기업등록부 회의와 2013년 UNECE/EUROSTAT/OECD 기업등록부 전문가그룹 합동회의에서 검토됨
- 가이드라인은 1장에서 SBR의 개념 및 정의, 활용, 가이드라인의 윤곽, 2장은 8대 역할, 조사틀, 조사지원, 정보자원, 국제자료 교환, 통계생산 및 서비스 현대화, 3장은 주요개념, 생산경제, 시장-비시장, 비공식부문, 제도단위 유형별 범위, 4장은 기업집단, 기업체, 사업체/지역활동단위, 지역단위, 특별목적단위를, 5장은 단위종류별, 인식 및 연락처, 생멸, 경제/계층 특성, 단위별 연계와 관계, 기타 등록 및 자료원, 6장은 행정자료, 통계자료원 및 통계조사 환류, 상업 및 기타자료원을 수록
- 7장에서는 SBR 유지관리, 생멸사건, 통계단위간 연계 혹은 통계단위의 존재에 영향을 주는 사건 예로, 생 혹은 멸, 합병 혹은 인수에 의한 집중, 본사 및 분할에 의한 비집중, SBR 내에 통계단위에 대한 연속성 규칙을 정의, 기업체는 만약 생산요소(토지, 노동, 자본)들이 계속되는 한 하나의 생멸사건을 통해 지속될 것 간주, 통계법정단위, 경제활동, 장소 3가지 특성 중 2가지 이상이 동일하게 유지하되 본질적으로 동일하다고 규정함. 9장 경제통계의 제공, SBR품질 충실성에 따른 경제통계 직접생산, 기업생멸통계 제공 등 11장은 SBR 구축시 핵심고려사항으로 SBR의 점진적 개발 인정 및 미래요구 모듈러 접근법 적용, 거버넌스 및 구성, 이용자들과의 관계, 법률체계, 자료원을 고려, 가이드라인의 적용 등을 수록하였음
- BR의 역할은 표본틀로서 표본조사 범주내에 통계단위의 총합이며, 총화·추출·연락에 필요한 단위의 특성을 포괄. 조사지원기능으로 자료수집(관측단위 정의 및 단위보고, 표본조사의 조정), 조사등록(단위의 선호되는 응답유형 보고서 기록), 조사통제(응답부담 측정 및 모니터링)실시. 정보원으로서의 SBR은 공개자료로서의 마이크로데이터 제공, 정부기관 및 기업을 위한 기업체 목록, 지리위치 정보, 소유 및 지분관리 정보, 정보프로그램을 지원하기 위한 활동 혹은 특정부문에 대한 정보 기밀보호를 준수해야 함. 통계 생산 및 서비스의 현대화를 위해 통계생산의 통합, 다른 자료원의 단위 연계 및 조정, 행정등록자료의 더 나은 활용, 행정등록자료와 전통통계조사자료의 결합, 통합자료를 제공하는 데이터웨어 하우스 제공기능도 수행함.

- 특히 SBR의 범주는 원칙적으로 GDP에 기여하는 국가경제내의 모든 단위여야 하고, 이는 2008 SNA 생산 범위(2008 SNA. 6.26)임. 즉 경제적 생산을 하는 모든 제도단위들의 생산 활동이지만 자가활동을 위한 서비스생산하는 가계 활동은 제외됨(단 자가정규거주서비스와 내부직원고용을 통해 생산된 서비스는 예외임). 시장 및 비시장 생산자도 포함됨
- 2016년 9월 가이드라인의 출간 1년 후에 기업등록에 관한 비스바덴그룹 및 유업통계위원회 추가의 대상으로 가이드라인이 어떻게 받아들여지는지를 설문조사를 실시하였고, 32개 국가에서 응답하였음. 21개 유럽, 아시아 6개, 아메리카 3개, 아프리카 1개, 오세아니아 1개가 발표됨. 질문 내용은 가이드라인의 홍보, 이행, 향후계획 및 기타 논평에 관한 사항이었고 이를 바탕으로 통계단위 및 프로파일링 강화, 모집단 기능 증대, 행정자료원 활용, 신자료원의 추가, 국제교육 및 세계화 추세 반영, 기업생멸통계 개발 등 개선사항을 발굴하였음.
- 아울러 아프리카 개발은행과 아시아 개발은행에서도 국제기구의 가이드라인에 근거하여 자체 기준을 마련하고 실제 이행사례 프로젝트였던 모리셔서통계청 BR 구축, 캄보디아-라오스 등 동남아시아 BR 구축 지원 프로그램 등에 대해서도 논의가 있었음.

□ 시사점

- 기업등록부의 가이드라인, 이행, 지속적 개선은 국제적인 공동 관심사안으로 한국형 통계 기업등록부 역시 이러한 국제기구들의 움직임을 상세히 파악하고 국제비교가능한 통계 기업등록부가 될 수 있도록 하여야 할 것임
- 기업등록부에 관한 비스바덴 국제회의 유치 혹은 국제기구 및 선진국 전문가 초청 세미나 개최 등을 통해 BR의 활용가능성을 높이고, 한국형 기업등록부가 국제표준을 선도해 나갈 수 있는 기회의 장을 마련해 보는 것도 바람직해 보임

[(IPS 060) 주목을 받고있는 데이터 사용자-통계 활용]

□ 세션개요

- 일시/장소 : 2017. 07. 19. 14:00~15:40 / A.1.02
- 세션구성
 - Organizer/Chair : Reija Helenius
 - Presenter ① : Elaine O'Mahoney
"Focus on users to increase the value of official statistics"
 - Presenter ② : Hannels Orjala
" The Customer Knows it better ?"
 - Presenter ③ : Emanuele Raldacci
"Putting users at the center : the European statistical system's approach in the DIGICOM project"
- 참석인원 : 약 45명

□ 세션요지

- 통계의 중심을 이용자관점으로 옮겨놓아야 한다는 점에 주안을 둔 세션으로 아일랜드 통계청, 핀란드통계청, 유럽통계청에서 추진하고 있는 이용자 중심 서비스 전략에 대한 다양한 논의가 있었음
- 아일랜드 통계청은 고객조사를 실시하였음. 응답자 규모는 506명이었고, 남성이 58% 여성은 42%였음. 81%가 35세 이상이었고, 77%는 상용종사자였음. 정치 및 공공기관 근무자가 40%, 금융 19%, 교육 15%, 과학기술엔지니어링 종사자가 11%를 차지하였음.
- 조사결과 매우불만족도 현황을 살펴보면 신뢰성이 3%, 정치개입 자율성 3%, 정확도 4%, 고객요구만족도 5%, 명확한 설명 8%, 이해편리성 7%, 서류명세성 10%로 나타났음, 이러한 응답결과를 바탕으로 해서 적정성 향상, 고객지향, 플랫폼 신설, 파트너십 구축, 교육기회 부여, 미디어와의 관계 개선 등의 다양한 고객중심 전략을 통해 개선방안을 구축하였음.
- 핀란드 통계청은 거리를 걷는 고객이 3D 안경을 통해 인구통계, 여행통계, 주택통계, 보조금 통계를 가상현실로 볼 수 있는 비주얼 통계전략 영상물을 제 1면에 보여주면서 발표장의 시선을 끌어드림.



고객의 종류를 이용자와 이해관계자, 응답자 혹은 자료제공자로 분류한 점도 눈여겨 볼 점. 분류된 고객에 맞추어 고객의 의견 환류를 받기위해 고객환류 모니터링 및 수집 연간 시계제도를 도입함. 특히, 고객의 관점으로 보는 통계는 신뢰성도 높아야지만 속도 역시 뒤쳐져서는 안 된다는 점임.



핀란드 통계청은 고객이 조직전략속에 보여야 하고, 통계개발과정에 고객과 이해관계자가 참여되어야 하고, 전체적인 고객의 평가 및 환류시스템이 조직운영관리에 포함되어야 함을 강조함.

- 유럽통계청의 경우에도 디지털의사소통 프로젝트(DIGICOM)의 일환으로 유럽통계시스템(ESS)이 고객중심으로 설정되어야 함을 피력하였음. 디지콤 프로젝트는 이용자 지식 개선, 이용자와의 관계강화, 신규이용자에게 다가가기, 이용자 관점의 프로젝트 진행으로 구성 되어 있음. 이를 위해 이용자심층 분석, 이용자 피라미드 구성(EU 제도 기관 등 Clinets, 통계청·은행 등 파트너, 언론·연구소 등 Heavy Users/ 공공정부·학생·교사 등 Intermediate Users / 정당, 개인 이용자 등 Light Users)로 분류하였음.



특히, 개별고객에 대한 가상이미지작성, facebook 개설, 시범통계(Experimental Statistics) 시연, 토론플랫폼 구성, 책자발간, 통계문맹 퇴치, ESO(European Statistics Olympics) 등 다양한 사업을 통해 고객에게 다가가려는 노력을 경주함.



□ 시사점

- 유럽통계처 및 유럽소재 통계청의 경우 이용자 중심의 통계서비스 전략수립에 분주하고 있음을 볼 수 있음. 고객들의 목소리를 그대로 반영하여 서비스를 개선하고 이를 통해 만족도를 제고시키는 모습이 상당히 인상적임
- 유럽통계처의 경우에도 유럽통계올림픽 등 이벤트 행사를 통해 이용자의 체험 기회를 확대하고자 하는 아이디어 등을 적극 발굴하는 모습을 벤치마킹할 필요성이 있음

[(IPS 104) 공연하는 통계학자가 된다는 것]

□ 세션개요

- 일시/장소 : 7. 18 / A.0.01
- 세션구성
 - 좌장 : Fabrizio Ruggeri
 - 발표 : David Spiegelhalter
- 참석인원 : 100명

□ 세션요지

- 발표자인 David Spiegelhalter는 캠브리지 교수로서 다양한 매체와 딱딱한 통계를 쉽고 재미있는 콘텐츠로 만들어서 전달하는 다양한 시도들을 보여주고 있음. 자칫 딱딱해서 통계를 싫어하는 이용자들에게 음악, 미술 등 문화콘텐츠처럼 통계도 대중화 되었으면 하는 바람을 강연 내내 표명하였음.
- David교수는 방송사에서 수영장을 배경으로 제작한 서바이벌 게임에 직접 출연하여 실패 및 성공을 통계로 수치화하여 시청자에게 몸으로 소통하는 통계 전달을 시도함. 아울러, 눈사람과 자신이 하늘을 나는 CG를 만들어 왜 미디어와 통계인이 만나야 하는지 다양한 설명을 시도. 우선 새로운 스토리에 대한 평가시도, 교육, 통계과학 로비, 연구와의 소통, 공공의 관심을 유도 하는 것 등이 그에 대한 이유라고 설명.
- 아울러, 일상 생활 뉴스로 볼 수 있는 미국인의 마리화나 흡연율, 지카바이러스 우려정도 등의 통계도 발췌하고 아울러 복권, 생일, 스포츠에 대한 재미있는 가십거리 통계도 콘텐츠로 발굴하여 대중과의 소통을 시도. 특히, 풋볼 우승에 대한 예측통계를 개와 경쟁하는 모습을 TV에 시연함으로써 통계로 웃음거리를 만들어 내기도 함. 아울러 국립왕실 통계협회의 통계대사로 미디어 훈련, 지역행사 참가 독려, 언론매체와의 연락 등 다양한 대외활동을 실시. 기후변화를 수치로 보여주는 BBC 다큐멘터리에도 출연하여 통계의 과학적 활용 홍보실시, 커피섭취에 따른 생명연장 통계, 생명표를 활용한 연령별사망위험률 통계 재산정, 스마트폰 활용이후 급격히 떨어진 성행위 통계 등 일반인이 쉽게 이해할 수 있는 자료들의 언론매체 노출로 통계의 활용도를 높임



- 이에 참석한 방청석에서도 많은 호응을 보이며, 춤과 음악을 겸비한 공연하는 통계교육 등을 통해 효과성을 본 사례들도 회자되었음.

□ 시사점

- 통계작성자 중심의 통계생산 패러다임에서 벗어나 이용자를 중심으로 통계가 만들어져 용이하게 사용되는 환경 인프라 조성을 강조되고 있음. 통계서비스를 중심으로 통계의 축이 변화되고 있음을 감지
- 이용자의 층을 다양하고 세밀하게 분석하고 계층별 촘촘한 이용자 만족 서비스 개발 및 통계이용 극대화를 위한 다양한 전략책 강구의 필요성 실감

[(IPS021) 국가중앙은행과 통계청 간의 통계협력 강화]

□ 세션개요

- 일시/장소 : ‘17.7.18(화) 10.30-12.30 / A.1.16
- 세션구성
 - 좌장(성명, 소속 등) Mr Chung Seak Roh(노 총식), 한국은행 경제통계국
 - 발표자(성명, 소속, 논문제목 등)

성명	논문제목
Mr Abdelhamid El Maazouzi	'Conditions for a successful statistical cooperation between national central banks and statistical offices'
Mr Joao Cadete De Matos	'Strategic coordination of a national statistical system: the case of the Portuguese statistical council'
Ms Begona Gutierrez Del Olmo	Holding companies and Head offices within the framework of the SNA 2008 ESA 2010: collaboration between the INE and the Banco de Espania for their sectoral classification and for obtaining data from annual business accounts
Mr Julio Santaella	Strengthening the national statistical system with the cooperation of the Central Bank: lessons from Mexico
Mr John Verrinder	Cooperation in Europe the macroeconomic imbalance procedure (MIP) scoreboard

- 참석인원 : 약60명

□ 세션요지

- 이 세션은 중앙은행 경제학자와 통계학자뿐만 아니라 중앙은행이 관심을 갖는 통계적 쟁점에 대해 토론하기를 원하는 포럼 기구인 IFC(living Fisher Committee on Central Bank)가 구성한 세션임
- 전 세계적으로 거시 경제 통계 자료는 여러 기관에서 제공하여 작성되는데 특히 각국 중앙은행(Central Bank, CB)과 국가 통계청 (NSO)이 생산하는 통계가 가장 많으므로 두 기관 사이에는 표준 및 방법론을 개발하기 위한 상당한 노력, 통계들 간의 일관성을 보장하고, 두 기관 간 협력 증대가 필요
- 성공적인 협력을 위한 세 가지 측면에서의 조건을 제시
 - (통계 거버넌스) 거버넌스는 통계 작성과 조직이 적절한 기능을 하도록 설정되어야 하며 통계 작성자들간에 제도적 협력을 공식화할 필요가 있음 특히 마이크로 데이터 측면에서는
 - i) 책임을 보다 잘 정의하고, ii) 중복을 피하며, iii) 응답자 부담 및 비용을 줄이고 iv) 효율적인 데이터 공유의 규칙을 정의
 - (통계생산 기관의 전문화) 일부 국가에서는 데이터 수집 시 중복을 발생하고 금융부문에 대한 중앙은행의 전문성과 실물 부문의 통계 기관의 전문성을 인정하지 않음에 따라 중앙은행과 통계기관 모두가 사업체조사를 실시하는 현상을 낳기도 하므로 각 기관의 전문성에 따라 조사를 분담할 필요 있음

- (기술 협력 및 전문 지식 공유) 거시 경제 통계의 생산을 조화시키기 위한 매뉴얼 개발 노력에도 불구하고, 여전히 용어 등에 대한 차이가 존재하므로 기술적 협력을 고취함으로써 차이를 줄여 나갈 필요

- ‘함께하게 되었다는 것은 시작(beginning)이고 함께를 유지한다는 것은 진보(progress)이며 함께 일을 해 나가는 것은 성공(success)이다’라는 Henry Ford(1863-1947)의 말을 인용(포르투갈 중앙은행)함으로써 성공적 국가통계시스템을 유지하기 위한 전략과 비전을 제시

□ 시사점

- 이 세션은 각국 중앙은행이 중심이 되어 구성된 세션이므로 중앙은행 입장에서 각국통계청과의 협력과 협조가 중요함을 제시하는 내용으로 구성되었으며 우리나라 한국은행 경제통계국 금융통계부 부장이 의장을 맡음
- 각국 중앙은행의 발표자중 Eurostat의 John Verrinder는 중앙은행과 각국통계청간의 관계를 ‘Wife and Husband’로 표현했는데 이는 ‘가깝고도 먼 사이’ 라는 것을 의미한다고 설명.
- 우리나라의 경우도 이런 가깝고도 먼 관계가 아닌가 싶었으며 그럼에도 불구하고 통계작성의 효율성이라는 측면에서 통계청과 한국은행이 서로 협력할 수밖에 없는 관계임을 인식하고 협력의 필요성을 인정해야 할 것임. 특히 GDP 중심 국민계정을 작성하는 한국은행과 지역 국민계정 (GRDP)을 작성하는 통계청이 상생의 관점에서 서로 협력할 필요성이 있다고 보임. 한편 이 세션은 각국 중앙은행 중심의 발표인데 각국 통계청도 함께 참여함으로써 서로간의 협력을 위한 상호의 관점에서의 발표와 토론이 이루어질 필요가 있다고 생각함

□ 첨부 및 사진 자료



(포르투갈 중앙은행 발표자: Joao Cadete De Matos의 발표내용중)



(chair 한국은행 금융통계부)

[(IPS078)설문조사 무응답 처리하는 혁신적인 접근법]

□ 세션개요

- 일시/장소 : ‘17.7.18(화) 14.00-15.40 / C.-1.01
- 세션구성
 - 좌장(성명,소속 등) Dr Linda Young/미국 농업부(U.S. Department of Agriculture)
 - 발표자(성명,소속,논문제목 등)

성명	소속	논문제목
Dr Jan Dempewolf	미국 메릴랜드 대학	Potential use of UAS systems for enumerating area frame segments for agricultural statistics
Dr Peter Vig Jensen	덴마크통계청	A holistic approach to eliminate the problem of nonresponse in surveys
Ms Barbara Rater	미국 농업부	Stemming the rising tide of nonresponse
Dr Michael Wild		Using real-time paradata analysis and response targeting in CAPI surveys

- 참석인원 : 60여명

□ 세션요지

- 가계 조사의 정확한 무응답률을 계산하기 위해 조사과정에서 수집된 실시간 paradata를 이용하는 방법을 제시함. 조사 관리자는 설문 조사 프로세스를 모니터링하고 응답자의 응답률이 전반적으로 낮아지는 경우 필요한 조치를 취할 수 있음. 파라데이터를 통해 수집된 자료의 품질을 평가할 수 있는 여러 리서치들이 있어왔고 세계은행의 ‘survey solution’ 이라는 파라데이터 수집 솔루션을 통해 응답률을 예측하고 목표 응답률을 높이는데 활용될 것이며 조사 과정의 효율성을 높여 응답률을 극대화 할 수 있을 것임. 그러나 현재 이러한 방법은 여전히 기술적인 문제, 법적인 문제 등이 놓여 있지만 관련된 문제를 하나씩 풀어나갈 계획
- 국가통계조사를 실시할 때 무응답을 줄이거나 없애기 위한 핵심적 방법으로 덴마크 통계청의 농업통계 작성 시의 사례를 발표함. 첫 번째는 행정자료를 이용하는 것인데 법률에 의해 덴마크 통계청은 정부기관에 의한 모든 등록 자료를 액세스 할 수 있는 권한을 갖게 됨으로써 농작물과 소에 관한 등록자료를 통해 통계를 작성할 수 있고 무응답문제를 해결할 수 있음. 행정자료는 기초적 자료제공뿐 아니라 자료를 유지 보완 관리를 용이하게 하는 유용한 자료가 될 수 있음

무응답을 줄이기 위해서는 주요 이해 관계자와의 협력이 중요한데 설문 조사를 수행 할 때 핵심 요소는 응답자가 참여할 의지이므로 응답자 친화적 조사표를 위해 설문 조사의 관련성에 대한 인식과 제공된 정보가 기밀로 유지 될 것이라는 신뢰가 필요. 따라서 덴마크 통계청은 농민 단체와 긴밀히 협력하며 이 협력의 결과로 농민 단체가 정기적으로 직접 농업

인들에게 덴마크 통계청이 실시한 조사에 참여하도록 권장하고 있으며 이 협조의 효과로 응답률에 긍정적인 영향을 미치는 것으로 나타남

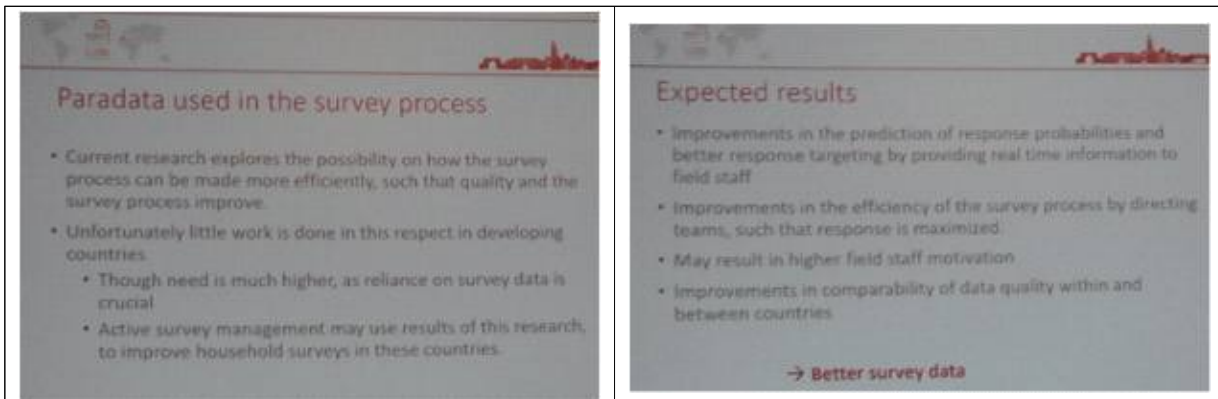
사용자 친화적인 데이터 수집을 위해 응답자와의 의사소통은 명확하고 이해하기 쉬워야함. 이를 위해 덴마크 통계청에서는 중앙 설문 조사 부서가 농부들에게 친숙한 표준화된 의사소통 도구 및 조사 설문지를 사용하여 응답자 친화적 조사표를 설계함. 이를 위해 의사소통은 이메일 또는 공개 시스템에 의해 수행되며 설문지는 질문을 웹 기반 및 대화형으로 조정하였으며 이 결과 응답률이 7% 증가하였음

□ 시사점

- 통계조사 과정에서 무응답을 줄이는 노력은 다양한 분야에서 시도되어 왔음. 우리나라도 현장 방문을 통한 조사자료를 행정자료로 대체함으로써 예산절감, 응답부담 감소 및 무응답 감소 등으로 통계의 내적 및 외부적 측면에서 품질과 효율성을 높여왔음.
- 조사과정에 관한 자료로서 파라데이터 수집 및 분석에 관한 논의는 우리청에서도 몇 년 전부터 시도되어 왔고 일부 모바일조사의 경우 별도로 조사하지 않아도 파라데이터가 수집 가능함. 그러나 좀 더 실질적인 활용 사례에 대해서는 아직 연구가 더 필요한 것으로 보임. 파라데이터도 일종의 빅 데이터의 일종으로 통계의 품질 요소 중 가장 중요한 요소라 할 수 있는 무응답률을 측정하는 데에도 활용 가능하며 특히 최근의 조사방법이 CAPI 또는 모바일로 확대되면서 수집은 더욱 용이해 질 수 있으므로 파라데이터의 실시간 수집을 통해 현장에서의 무응답을 줄일 수 있는 정보제공에 많은 활용이 있을 것으로 기대됨
- 응답자의 인지를 실험하여 설계과정에 반영하는 우리청의 조사표 실험과정도 이러한 응답자 친화적인 조사표를 설계하여 무응답을 줄이기 위한 노력으로 볼 수 있음. 또한 응답자 간담회 등과 같은 응답자와의 협력 증진 활동도 동일한 노력이라 할 수 있음

□ 첨부 및 사진 자료

<조사과정에서 Paradata의 활용 및 기대효과>



[(IPS053)공식통계의 생산을 별도의 정부부서에서 할 필요가 있는지 혹은 정부부서내에 남아 있어야 하는가?]

□ 세션개요

- 일시/장소 : '17.7.19(수) 10.30-12.30 / A.1.18
- 세션구성
 - 좌장(성명,소속 등) Mr Hallgrimur Snorrason
 - 발표자(성명,소속,논문제목 등)

발표자	논문제목
Mr Pali Lehohla	Statistics in politics: from skiing on the broad and smooth to the serrated side of the saw
Dr Andreas Georgiou	The production of official statistics needs to be a separate branch of government
Mr Jean-Luc Tavernier	A statistical system integrated in the national central administration: the French experience
Mr Dennis Trewin	What does an independent statistical agency mean in practice?

- 토론자(성명,소속 등) Mr Gerry O' Hanlon
- 참석인원 : 약 100명

□ 세션요지


- 전 호주 통계청장인 Dennis Trewin이 ‘독립적 공식통계작성기관이 실질적으로 의미하는 것은 무엇인가?’ 라는 제목의 발표에서 통계기관에 있어 가장 기본은 **신뢰성**임을 강조하였고 통계의 신뢰성이란 ‘통계는 객관적으로 작성되고 발표되어야 하며 통계의 품질에 확신이 있어야 하고 공식통계의 사용은 자격요건이나 질문 없이 사용되어야 하며, 비밀보호는 통계의 신뢰를 유지하기 위한 필수 요건’임을 강조함
- ‘독립적 공식통계작성기관이 실질적으로 의미하는 것은 무엇인가?’에 대해서는 구체적으로 정부와 핵심 이용자들 간 긴밀한 관계 형성(독립성이란 정부와 주요 이용자의 요구와 동떨어져 있다는 것은 아님), 통계적 방법, 발간시기와 발간 내용을 결정하는데 있어 독립성 유지, 발간일이 변경 되어야 할 때는 공적인 설명을 통해 가능한 발표 일정을 공지하며, 매체들과 강한 우호적 관계 및 독립적 통계기관의 중요성에 대한 대중의 이해와 수용이 있어야 함을 강조.
- 끝으로, 공식통계는 사회에 대한 거울(mirror)을 제공하며. 통계는 정책입안을 할 때, 특히 민주사회에서 중요한 역할을 함. 따라서 공식통계에 있어 신뢰는 필수 불가결하고 신뢰야

말로 실질적으로 독립적 기능을 가능하게 하는 요소임을 강조. 그 외에 프랑스의 INSEE를 중심으로 한 16개 통계작성기관이 중앙정부체계 내에서 통합된 통계 시스템을 갖고 있으며 INSEE의 정치적 간섭으로부터 완전히 독립적인 체계(엄격한 엠바고를 통해 통계를 독자적으로 보급하고, 언론에 발표)를 유지하는 사례를 발표

□ 시사점

- 발표하는 연사 중 2004년에 우리나라 통계청을 방문했던 당시 호주 통계청장이었던 Dennis Trewin인 것이 무척 인상적이었고 공식통계를 작성하는 기관의 핵심요건은 신뢰성 (Trust)임을 다시 한 번 느낄 수 있는 세션이었음.
- 통계작성기관이 어떠한 이유로 신뢰성을 잃는다면 그 기관에서 작성되는 통계는 더 이상 믿을 수 없는 수치에 불과할 것이며 신뢰할 만하고 타당한 통계를 작성하기 위해 정부기관은 법률적 프레임웍과 자원을 통계 시스템 내에 통계작성기관에 제공해서 통계인(통계작성자)이 최선의 가능한 방법과 기술을 사용하는데 있어 간섭이 없이 요구되는 통계정보를 생산하도록 해야 할 것 이라고 점을 강조
(Dennis Trewin 발표자료 중, Malaguetta and MacDonald, J of IAOS, 2016)

□ 첨부 및 사진자료

	<p>Requirement</p> <p>"To produce reliable and valid statistics it is necessary that Governments provide the legal framework and resources to the statistical systems of their countries to allow statisticians to produce the required statistical information, without interference using the best available methodology and techniques from the best suited sources of inforamtion." (Malaguetta and MacDonald, J of IAOS, 2016)</p>
<p>(발표 연사, 맨 왼쪽 남아공 전 통계청장(Pali Lehohla), 세 번째 호주 전 통계청장, Dennis Trewin)</p>	<p>(Dennis Trewin 발표자료 중, '신뢰성 있는 통계작성을 위해 요구 되는 것')</p>

[(IPS136) 빅데이터와 큰 기회들]

□ 세션개요

- 일시/장소 : '17. 7. 18 (화) 10.30-12.30, A.0.01
- 세션구성
 - 좌장 : Thuan Nguyen
 - 발표자
 - ① Danny Pfeffermann(이스라엘 통계청: Can Big Data Really Replace Traditional Surveys for the Production of Official Statistics)
 - ② Cinzia Cirillo(미국 메릴랜드 대학교: Prediction of traffic patterns and travel time using Big Data: a small area estimation approach)
 - ③ Monica Pratesi(이탈리아 통계청: Small Area model for Italian regional monthly estimates of young unemployed using Google Trends Data)
- 참석인원 : 약 50여명

□ 세션요지

① 공식 통계작성에서 빅데이터를 활용한 전통적 조사방식 대체

(Can Big Data Really Replace Traditional Surveys for the Production of Official Statistics)

- 빅데이터로의 접근 및 분석기술 발전, 데이터의 정확성 요구, 긴축예산 등 빅데이터를 활용하여 기존 표본조사를 대체하기 위한 불가피한 압력이 가중되고 있음
본 발표에서는 공식통계를 위해 **빅데이터를 사용하는 중요한 도전과제를 논의하고, 그 장점과 한계**에 대해 언급함
- 두 번째로, 제한된 **인구데이터와 빅데이터를 연계하는 통계모델을 제시**
- **빅데이터 관련 주요 이슈**
 - (**Coverage**) 대표성은 빅데이터를 공식통계화 하는데 가장 문제가 되는 부분임(예: SNS에 의견을 피력하는 이용자들이 전체 국민을 대표하지 못하는점)
 - (**Bias**) 빅데이터가 다른 변수의 예측변수로 쓰일때는 편향이 없음
 - (**Sampling**) 무작위 추출은 빅데이터 분야에서 여전히 주요한 역할을 하는데, 제한된 모수 내에서 랜덤추출과 빅데이터에서의 랜덤추출은 다르므로, 추가 연구 필요
 - 하위 인구그룹을 위한 빅데이터: 각국 통계청은 하위 인구그룹의 추정치를 발표하는데 빅데이터는 나이, 성별, 국적, 지리 등의 정보를 포함할 수 없음(예를 들면 슈퍼마켓에서 사용한 신용카드의 사용자의 개인정보는 포함하지 않음)
 - 경험적 우도 접근방법: 비모수적 인구분포와 빅데이터를 위한 모수적 모델을 결합하고, 보정 제약조건을 추가함

- 빅데이터와 조사자료 결합

(가정): 빅데이터에서의 표본 요소는 정해져 있음

- 베이저안 정리: 빅데이터는 대표도 편의, 응답편의 또는 측정오차 문제가 있으며, 이러한 비표본 오차 문제를 해결하기 위해 빅데이터 추론을 위한 베이저안 정리를 이용

* Y는 c(확률변수)에 대해 표본멤버십을 나타내는 조건변수이며, 빅데이터는 대상 모집단의 샘플이라고 가정

빅데이터와 조사자료 결합	베이저안 정리
<p style="text-align: center;"><u>Survey integration, combine big data with survey data</u></p> <p>Basic assumption: Membership of sample elements in big data (B) set known. (Ask the sample members?)</p> <p>Let $A = \{i i \in B\}$ (i.e. B). Denote the target variable by Y.</p> <p>Sample data: $\{(x_i, z_i, \beta_i) i = 1, \dots, n\}$.</p> <p>$x_i$ = model covariates, z_i = variables explaining B-membership</p> <p>Procedure: Model $\pi = \text{Pr}(i \in A x_i, z_i)$ from sample data $\Rightarrow \hat{\pi}$</p> <p>Use $w_i = (1/\hat{\pi}_i)$ as weights for analyzing big data.</p>	<p style="text-align: center;"><u>Bayes Theorem</u></p> <p>For Y, C random variables.</p> $\text{Pr}(Y=y C=c) = \frac{\text{Pr}(C=c Y=y) \times \text{Pr}(Y=y)}{\sum_j \text{Pr}(C=c Y=y_j) \times \text{Pr}(Y=y_j)}$ $f_Y(y C=c) = \frac{\text{Pr}(C=c Y=y) f_Y(y)}{\text{Pr}(C=c)} = \frac{\text{Pr}(C=c Y=y) f_Y(y)}{\int \text{Pr}(C=c Y=y) f_Y(y) dy}$ <p>C = conditioning variable indicating sample membership.</p> <p>◆ In what follows I assume that the big data are a "sample" from the target population</p>

② 빅데이터를 이용한 교통패턴 및 여행 이동시간 예측 : 소지역 추정 접근법

(Prediction of Traffic Patterns and Travel Time using Big Data: A Small Area Estimation Approach)

- ATIS(Advanced Traveler Information System)는 **여행 전(前)과 여행 중 사용자에게 충분한 정보를 제공**하여 의사결정을 내릴 수 있게 해 줌, 동적 경로안내 시스템은 ATIS를 사용하여, 덜 혼잡한 경로를 따라 자동으로 경로를 지정하고, 예상 도착시간을 지속적으로 통보함. 또한 실시간 데이터를 지속적으로 수집, 분석하여 트래픽 관리 및 작동을 향상시키고, 최신 정보를 제공함

- 그러나 실시간으로 수 천개의 도로구간을 예측하는 경우, 방법론 및 계산방식상 문제가 발생, 본 발표에서는 **소규모 지역 추정(SAE)의 아이디어를 차용**하는 새로운 방법을 제안하고, **소지역 추정치가 자동차용 프로브 프로젝트(VPP:메릴랜드 대학 운영) 빅데이터를 활용하여 실시간 교통예측을 실현할 수 있는지에 대해 설명**

- 교통관련 문헌과 경험적인 기존 지식에 의하면, Y(교통속도)는 t(시간)와 w(주)가 고정된 경우에도 d(요일)에 따라 달라짐. 전통적 예측방법에 비해 제안된 내용의 새로운 점은 (1) 예측 전단계에서 모델 선택/ 진단, 피팅 및 유효성 검사를 수행할 수 있기 때문에, 수천개의 도로상황을 즉각적으로 예측 가능

(2) 종합적인 가정에서, 부분이 아닌 모든 데이터를 사용하여 보다 유익한 모델 사용

③ 이탈리아 실업률을 이용한 소지역 모델 기반 추정 : Google 트렌드 데이터

- 많은 연구자들은 공식 경제지표를 개선하기 위해 온라인 인터넷 데이터 사용을 제안함. 최근 문헌에 따르면, **구글 트렌드 구인광고 시리즈**를 활용한 이탈리아 실업률 관련한 다양한 시계열 모델이 제시됨. 이 논문의 목적은 **인터넷 검색 데이터 사용을 이탈리아의 소규모 지역실업률 추정까지 확장**하는 것임

□ 시사점

- 빅데이터의 공식통계 활용은 대표성 부족 등의 품질문제, 모집단의 부재로 인한 추정치 작성 불가 등의 한계가 있으나, 이를 해결하기 위해 각국에서는 이론적, 제도적 노력이 다양하게 진행되고 있음
- 자료수집 비용 및 응답부담 증가, 시의성 있는 신규통계 수요 증가 등 통계작성 환경의 변화로 인한 새로운 자료수집 방법 필요에 따라, 공식통계 작성기관에서는 빅데이터 통계를 공식통계로 활용하기 위해 새로운 통계작성 패러다임으로의 변화를 도모해야함

□ 발표 관련 사진



[(IPS012) 최근 발전되고 있는 빅데이터 공동모델]

□ 세션개요

- 일시/장소 : '17. 7. 17 (월) 14.00-15.40, A.0.01
- 세션구성
 - 좌장 : Thaddeus Tarpey
 - 발표자
 - ① Weiliang Qiu(하버드 의과대학 보스턴 메디컬 스쿨: Joint tests for equality of mean and variance with application in high-throughput DNA methylation data analysis)
 - ② Renjun Ma (캐나다 New Brunswick 대학 수리통계학과: Joint-modelling of discrete, continuous and semi-continuous longitudinal Big Data with applications)
 - ③ Lang Wu(Joint models for HIV vaccine data)
- 참석인원 : 약 30여명

□ 세션요지

① 대용량처리 DNA 메틸화 데이터 분석: 평균과 분산의 동일성 결합검증

(Joint tests for equality of mean and variance with application in high-throughput DNA methylation data analysis)

- DNA 메틸화는 유전자 코드 변경 없이 유전자 발현을 조절하는 유전적 메커니즘으로, 흡연 등 환경적 요인에 의해 수정될 수 있으며, 유전/환경적 요소 및 암과 같은 복잡한 질병에 있어 중요한 역할을 함. 특히 각 DNA 메틸화 지수에서, 질병에 감염된 대상과 건강한 대상의 평균 메틸화 수준에 대해 검증가능 (* 평균뿐 아니라, 분산도 통계적 분포를 결정하는 중요한 요소임)
- 최근 몇 연구원들이 DNA 메틸화 지수의 분산을 비교하는 것을 제안하였으며, Ahn과 Wang(2013년)은 질병에 걸린 대상과 건강한 대상 두 그룹간의 평균과 분산을 함께 비교하기 위한 결합 검증을 제안
본 논문에서 더욱 복잡한 질병의 DNA 메틸화 지수를 살펴보기 위해 새로운 joint test를 제안하고, 이는 더욱 복잡한 질병의 예방, 진단, 치료에 사용될 수 있을 것
- DNA 메틸화: DNA 메틸화는 DNA에 메틸 그룹을 추가하는 것을 포함하며, 특히 유전자 영역의 메틸화는 유전자 발현을 억제할수 있음. 메틸화는 일반적으로 CpG 뉴클레오타이드에서 발생함
 - 질병이 있는 피험자와 건강한 피검자간의 다른 DNA 메틸화 수준의 CpGsite를 가짐

- 차별 가변성 연구를 하는 이유: 최근 암 후성 유전학에서는 변동성 측면에서 다른 특징을 확인하는 것이 질병 표현형을 이해하거나 예측하는데 중요하기 때문
- 동등성 분산테스트 방법: Feinberg and Irizarry(2010) 등 연구자들은 전통적인 F 검정 방식으로 검증하였으며,
- 동등성 분산테스트 비교: conover et al.(1981)은 시뮬레이션을 통해 56가지 동등한 분산테스트를 비교하였는데, BF 방식이 적절한 검정력을 가지고, 명목상 오류를 유지하는 동등한 분산테스트 중 하나임을 발견함
- 결론: **정상 가정하에서 우도비 테스트가 가장 효율적**이었는데, 유형1 오류율을 0.05로 유지하면서, 최대 검정력을 가짐. 실제 데이터 연구에 따르면, 제안된 3가지 결합테스트 중 우도비 테스트가 가장 효과적이었음

② 이산, 연속, 반연속 종단 데이터의 결합 모델링 응용 프로그램

(Joint-modelling of discrete, continuous and semi-continuous longitudinal big data with applications: Renjun Ma)

- 이산, 연속, 반연속 종단 데이터는 생물학, 환경, 보건 사회연구에서 주로 다루는데, 본 연구에서는 분포가 없는 무작위 효과를 갖는 Tweeie 일반 선형 모델을 사용하여 혼합유형의 종단 데이터를 공동 모델링하는 것을 제안함. 이러한 모델에 대한 최적의 추적기능은 임의 효과 최고의 선형 비편향 예측을 기반으로 개발되었음
- (조랑말 근전성 실험) 6마리는 무작위로 플라시보 실험, 9마리는 약물투여
 - 양적 지속 반응을 살펴보고, 각 조랑말은 64회 실험을 하여, 종단데이터를 확보
- (보험자료 연구) 각 가입자의 납입금수 및 납입금액 총수
- HIV 동적 연구 및 허혈성 심장병 관련 연구 등 수행

③ HIV백신 연구를 위한 결합 모델

- 최근에 몇 년간 종단 및 생존 데이터에 대한 결합모델이 주목을 받고 있음. 본 발표에서는 HIV백신 연구를 위한 결합모델을 고려하는데, 다음과 같은 문제가 발생할 수 있음. (i)종단 데이터가 2진 및 연속 데이터와 같은 다양한 유형의 데이터가 혼합될 수 있음 (ii) 수량의 한계로 인해 종단 데이터를 검열 (iii) 종단 방향 데이터는 오류로 측정할 수 있으며 값 누락 가능 또한 우도비 추론과 관련된 계산은 매우 까다로움. 발표자료에서는 앞서 언급한 문제를 해결하기 위한 결합 모델과 효율적인 계산방안을 제시함

□ 시사점

- 의학, 생물학, 보건 사회분야에서의 빅데이터 분석은 기존의 평균 뿐 아니라 분산까지 비교함으로써, 질병의 예방, 진단, 치료에 효과적인 방안을 모색함. 또한 결합 모델을 통해 기존의 연구를 보완하고, 효율적인 산식을 제시가능함

□ 발표 관련 사진



[(IPS024)빅데이터시대의 공식통계]

□ 세션개요

- 일시/장소 : '17. 7. 18 (화) 16.00-17.40, A.0.01
- 세션구성
 - 좌장 : Aurel Schubert
 - 발표자
 - ① Adopting Big Data technologies in the production of official statistics: opportunities, experiences and lessons learned (이탈리아 통계청, Antonino Virgillito)
 - ② Assessing the use of Google Trends to predict credit developments Edwige Burdeau
Improving financial information processing: a case for a Big Data architecture(프랑스 중앙은행)
- 참석인원 : 약 50여명

□ 세션요지

① 공식통계 작성을 위한 빅데이터 기술채택: 기회, 경험 및 교훈

(Adopting Big Data technologies in the production of official statistics: opportunities, experiences and lessons learned)

- 공식통계 작성을 위한 빅데이터의 사용은 최근 몇 년간 국제 수준에서 여러 자료원에 대한 접근, 새로운 방법론개발 및 새로운 도구사용과 같은 측면에 중점을 둠. 그러나 병렬 컴퓨팅 플랫폼 및 NoSQL 데이터베이스와 같은 기술통계, 각국 통계청에는 기술이 빅데이터 혁명의 주요 원인이었음에도 불구하고, 주로 실험활동으로 제한됨. 본 발표에서는 빅데이터 기술을 활용한 통계작성 지원 및 이탈리아 통계청 및 국제수준에서의 경험을 분석하는 다양한 방법을 제시함.
- 또한 빅데이터의 도입이 통계전문가의 업무흐름과 IT 부문의 역할에서 어떤 중요한 변화를 필요로 하는지 설명하며, 기술의 잠재력을 최대한 효과적으로 활용하려면 어떻게 데이터 흐름을 합리적으로 재구성해야하는지 논의함

* **UNECE의 Sandbox:** 공식통계 생산을 지원하기 위해 빅데이터 활용방법 공유하기 위해, UNECE(유럽경제위원회) HLG가 빅데이터 프로젝트의 일환으로 구축한 빅데이터 플랫폼으로, 최근에는 유럽통계청(Eurostat)의 빅데이터 프로젝트를 위한 환경지원 및 빅데이터 교육을 위한 플랫폼 등으로 활용

** **이탈리아 통계청의 빅데이터 플랫폼:** 자체 개발한 빅데이터 플랫폼으로, 8-node 하둡 클러스터+spark+MPP DB, 통계생산 및 실험이 가능하도록 설계됨 (스캐너 데이터(5-7천만건/월) 및 모바일폰을 활용한 인구추정 등)

- 크기로 인한 문제: 데이터의 양은 빅데이터에서 실질적으로 가장 큰 영향을 미치는 요소이며, 절대 임계점이 없다는 것, 데이터를 이동시키기 어렵다는 점, 자료처리 시간이 오래 걸리며, 통계전문가들이 접근하기 어렵다는 문제가 있음
- 빅데이터의 증대는 통계분석과 IT 기술의 교차점에 있는 통계조직에 새로운 문제를 야기 시키는데, 통계비즈니스 측면이라는 페러다임의 전환이 요구되며, 공식통계 내부 이용자를 위한 데이터 통합 온톨로지 및 외부이용자를 위한 공개데이터 온톨로지가 필요함

② 신용평가를 예측하기 위한 구글 트렌드 사용 평가

- 예측에 관한 자료들은 보통 구글 검색 데이터를 통해 이루어지고 있는데, 특정기간 동안 구글에서 검색한 키워드는 웹 사이트 구글 트렌드를 통해 실시간으로 무료로 제공. 또한 Google Correlate도 주목을 받고 있는데, 이는 관심있는 사항에 대해 많은 잠재 예측 변수를 사용할 수 있는 장점이 있음
- 본 발표에서 주택구매를 위한 프랑스 신용 흐름을 예측하기 위해, 구글 트렌드 및 구글 correlate가 유용한지 알아봄
- 구글 트렌드와 관심변수간의 관계가 반드시 선형적이지는 않으므로, 이러한 지표의 예측력은 기계학습방법으로 향상 가능하며, 따라서 LASSO(Least Absolute Shrinkage and selection operator) 및 BMA(Bayesian Model Averaging)과 같은 선형방법뿐 아니라 부스팅 모델 및 Support Vector Machine과 같은 비선형모델도 포함
- **(구글 트렌드)** 경제의 수요측면을 파악하는데 유용하며, 본 연구에서 주택 지표보다 0-3개월 후 예측: 구글 트렌드에서의 설명지표로 구글 correlate에서 적용시도
- **(구글 correlate)** 자동 쿼리를 생성하는 두가지 방법: 주택 구매를 위한 크레딧으로 외부 시계열을 업로드하면 구글 correlate가 가장 연관성이 높은 구글 트렌드를 알려 주고, 프랑스 신용 기관의 이름이 포함된 12개 지표를 수동으로 선택-> 미리 지정된 개별 구글 검색어(주택 구매에 대한 신용)을 제공함으로써, 구글 correlate는 가장 높은 상관관계가 있는 100개의 검색어를 반환함
- **(기계학습 모델의 유용성 실험)** 미리 선택된 관련없는 지표의 노이즈를 제거하기 위한 다양한 변수선택 모델(LASSO(Least Absolute Shrinkage and selection operator): 관련 없는 변수의 값은 0으로 축소됨
 - * 베이저안 모델: 관련없는 변수를 포함할 사후확률이 낮음
- 결론: 구글 트렌드와 구글 correlate를 동시에 사용하면 중기적 예측을 크게 개선 가능하며, 주택구매를 위한 신용흐름의 선행지표를 확인함

□ 시사점

- 빅데이터 관련 통계조직 내에서 IT 전문가와 통계전문가 관계도 중요한 문제로, 서로 배제하지 않는 환경에서 각자의 역량강화가 중요함

□ 주요 발표 사진

공식통계 작성을 위한 빅데이터 기술채택 발표	구글 트렌드 예측
	