

제4차 PAPOR 연례회의 참가 결과 보고

2015. 12.

통계개발원
동향분석실
박 시 내

차 례

I. 회의 및 출장개요	1
II. 주요내용	3
1. 단기코스: Dual Frame Cell Phone & Landline Surveys: Design and Weighting	3
2. 일반 세션	6
가. 모바일 조사의 혁신과 동향	6
나. 미국 Census 자료에서 Race와 Ethnicity의 측정	9
다. Beijing Consensus 모델 검증	11
라. Big Data 활용: Text & Sentiment 분석	13
III. 참가소감	15

I. 회의개요

1. 참가회의 : PAPOR* 연례회의

* Pacific Chapter of American Association for Public Opinion Research Annual Conference
 **'15년 논제: "Measuring Public Opinion in an Expanding Opinion Landscape"

- 주 관 : AAPOR(미국 여론조사 학회)

- ◆ AAPOR(미국 여론조사 학회)은 조사방법론 분야에 있어서 미국 내 전문가 뿐만 아니라 전 세계 전문가들의 네트워크 단체
- ◆ 1947년 설립되었으며 1,900여개의 학계, 비영리단체, 여론조사기관, 정부 내 통계 전문가, 통계학자 등으로 구성
- ◆ 조사연구 전문가들의 네트워크 형성 등을 위해 매년 5월 연례회의를 개최하고 있으며, '12년부터 매년 12월 태평양 지부 연례회의를 개최함

- 회의기간 : 2015. 12. 10(목)~11(금) / 2일

- 장 소 : 미국 샌프란시스코, Marines' memorial Hotel*
 * 609 sutter st, San Francisco, CA 94102

< 컨퍼런스 일정 >

일자	세션	주요주제
12.10(목) 7:30 am	등록	
12.10(목) 8:00 ~ 12:45	Short Course 1	· Dual Frame Cell Phone and Landline Surveys: Design and Weighting
	Short Course 2	· Mobile Devices and Modular Survey Design
12.10(목) 오후	Session A	· 표본, 디자인, 분석방법
	Session B	· 조사응답과 측정개선을 위한 방법/ 빅 데이터
12.11(금)	Session C	· Survey의 혁신적인 방법과 접근방향
	Session D	· 여론 측정의 혁신과 동향
	Session E	· 모바일 조사의 응답과 설계

2. 출장개요

- 출 장 자 : 동향분석실 박시내 사무관

- 출장기간 : 2015. 12. 9(수) - 13(일) / 5일

< 출장 일정 >

월 일 (요일)	출발지	도착지	방문기관	업무수행내용	접촉인물 (소속포함)
5.14(수)	인천	샌프란시스코	-	출국 및 이동	-
5.15(목)			샌프란시스코(Marines' memorial Hotel)	등록 및 단기코스, 세션(A,B) 참가	Sarah Cho (Survey Monkey)
5.16(금)			"	세션(C,D,E) 참가	Tom Wells (Nielsen Company)
5.17(토) ~ 5.18(일)	샌프란시스코	인천		귀국	-

3. 수행사항

- (조사방법론 분야의 최근 동향 파악) 모바일 조사, 웹 조사 등 조사기법의 최근 동향 파악 및 표본추출, 자료 분석 관련 최근 동향 파악 및 자료 수집으로 연구 추진*에 기여
 * 순환센서스 2차 시험조사 혼합조사 도입 방법 연구
 **미국 센서스 국 인종의 측정 및 활용 등 관련 연구 참고 예정(순환 조사표 설계)
- (관련 연구자와 인적교류) 민·관·학의 통계조사방법 분야의 다양한 전문가들과 인적 네트워크 형성 및 연구 성과 공유

< 컨퍼런스 전경 >



II. 주요내용

1. Short Course

단기
코스

"Dual Frame Cell Phone & Landline Surveys: Design & Weighting"

12.10(목) 8:00 a.m. - 10:15 a.m

Presenter : Stas Kolenikov, *Abt SRBI*
e-mail : S.Kolenikov@srb.com

- Cell phone survey는 2000년대 중반에 새롭게 대두된 조사방법으로, 셀폰이 널리 보급되면서, 용이한 조사방법으로 인식되나 응답률이 낮다는 특성을 가짐
- 특히 RDD(random digit dialing) Cell phone 표본이 RDD 유선전화 샘플과 결합될 때 표본에 대한 가중치 부여방식에 대해서는 합의되지 않음*
 - * AAPOR TF on Future of Phone Research Report(2010)

Official (dated?) statement from AAPOR

"There is no consensus regarding how RDD cell phone samples should be weighted, especially when combining them with RDD landline samples."

- 2010 AAPOR Cell Phone Survey Task Force Report

Watch out for AAPOR TF on Future of Phone Research Report

Introduction

While there is no consensus approach to weighting, a number of the methods available work quite well

DSR rank	Pollster	# of Polls	NCFP / AAPOR / Roper	Calls / Cellphones?	Predictive Plus-Minus	DSR Grade
1	SurveyUSA	72	✓	Sometimes	-0.9	A-
17	CHN/Ogden Research Corp.	72	✓	✓	-0.9	A-
28	Marist College	69	✓	✓	-0.9	B+
30	Quinnipiac University	133	✓	✓	-0.3	B+
35	Mass-Open Polling & Research, Inc.	407	✓	✓	-0.3	B+
51	University of New Hampshire	76	✓	✓	-0.1	B-
69	Public Policy Polling	297	✓	✓	+0.0	B-
92	YouGov	230	✓	✓	+0.3	C+
126	Strategic Vision, LLC	130	✓	✓	+0.3	F
128	Stratix	80	✓	✓	+0.4	C+
180	Research 2000	279	✓	✓	+0.6	F
212	Zogby Analytics, Telephone	379	✓	✓	+0.7	C
238	Research	623	✓	✓	+0.7	C
248	American Research Group	223	✓	✓	+0.8	C-
314	Survey Interactive, online	128	✓	✓	+1.1	D+
330	Zogby Interactive/J Analytics	38	✓	✓	+2.9	F
337	IC Research	133	✓	✓	+2.9	F

< 강연 모습 >



- Dual frame RDD 조사에 최적화 기법에 관한 연구 수행됨*

* Optimal allocation paper(Lohr & Brick 2014 JSSAM; Levine & Harter 2015 POQ) 등
** 다중표본추출틀 표본설계(multiple frame sampling designs)는 2개 이상의 표본추출틀을 함께 사용하는 표본설계방법으로 미국 센서스국은 1949년 소매상 조사에 처음으로 다중표본추출틀을 사용함

- 표본틀에 따라 under-coverage 및 over-coverage에 차이 있으며, 유선전화만 있는 경우 가구기준 8.4%, 셀 폰만 있는 경우는 45.4%, 둘다 있는 경우는 42.7%, 모두 없는 경우는 3.2% 수준임(2014년)

Frame imperfections

Frame	Undercoverage: eligible units missed	Overcoverage: ineligible units present on frame
Landline phone	Cell phone only No phone service	Inactive numbers Business numbers Fax machines, DSL, data
Cell phone	Landline only No phone service	Inactive numbers Business numbers SIM cards in remote sensors
Address based samples	Homeless	Vacant units Business addresses
Areal samples		Business addresses
Social networks	People not on the network	Abandoned accounts (?)

Dual frame RDD

- Domain a: landline only
 - 7.1% adults, 8.4% households nationwide
- Domain b: cell phone only
 - 44.1% adults, 45.4% households nationwide
- Domain ab: overlap, dual users
 - 45.8% adults, 42.7% households nationwide
- Domain o: no service
 - 2.9% adults, 3.2% households nationwide

Source: July-Dec 2014, NHIS (Blumberg and Luke 2015)

- (다중표본추출틀 조사에서 중복 문제의 해결 방법) ① Screening 방법, ② 표본의 추출확률을 이용하여 가중치로 보정, ③ 추정단계에서 중복된 부분을 사용하여 효율적인 추정량 얻는 방법

- (표본의 대체와 강화) 전화조사가 가능 혹은 불가능한 표본을 전화조사 전에 알 수 있는가? 이중 표본추출틀에서 응답자의 특별한 유형별 표본의 대체, 특별한 지역에 대한 샘플 확보가 가능한가?*

* Brick(1995) listed-assisted RDD 방법

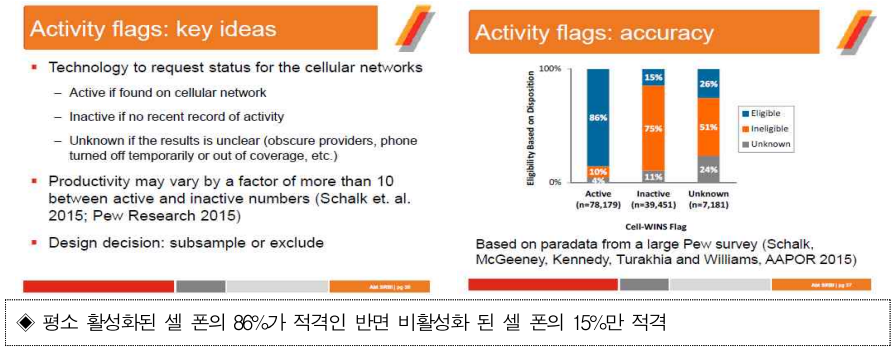
Geography

- Rate center
 - The rate center associated with the LL or CP number
 - Legacy of the era of expensive long distance calling
 - Area code + exchange + 1000 block
- Billing ZIP code
 - The ZIP code associated with the primary cell phone account holder (if available from the cell carrier)
 - Appended to ~40-60% of the sample
 - Extra product to be purchased on top of the sample
- Modality tract / county on the LL
 - Geographic unit where most listed numbers from bank are found

Geography

○ **(Landline frame)** 등재된 번호를 이용하여 모든 100-banks 중 가구용으로 사용될 수 있는 100-banks들을 모두 확인 후 특정 개수 이상의 등재번호들이 들어있는 100-banks에서 1단계 동일확률추출법을 이용하여 전화번호 추출함

○ **(Cell phone frame)** 7-digit 번호 조합의 1000-banks에서 표본을 추출하는 방식을 사용하여 RDD 표본 추출함. 셀 폰 프레임에서 중요한 것은 셀 폰 사용자의 특성이 고려되어야 함



○ **(가중치 부여 요소)** Baseline probability of selection, Eligibility adjustments, Multiplicity adjustment, Unit nonresponse adjustments, calibration

○ **(Dual frame RDD 가중치 작성 방식)** 응답자의 전화유형(유선 or 셀 폰), 지역, 전화기 개수 등을 이용하여 응답자의 추출 확률 산출, 응답자 추출 확률을 이용하여 전화 보유 상태의 세 가지 유형(① 유선만 보유, ② 셀 폰만 보유, ③ 유선 & 셀 폰 모두 보유)별 초기 가중치 산출(① 중복을 고려한 경우 ② 중복을 고려하지 않은 경우) 과정을 통해 총 6가지의 가중치 산출됨

○ **(Dual frame RDD 가중치 조정방법)** 가중치 조정 절차는 ① RDD 초기 가중치, ② 비적격 번호 제거, ③ 가구 무응답 조정, ④ 개인가중치, ⑤ 개인 무응답 조정, ⑥ 사후층화 방식으로 이뤄짐

○ **(결론 및 시사점)** 디자인, 표본, 가중치와 추정은 모두 연계되어 있으며, 유선전화와 셀 폰의 이중표본추출에서는 여러 가지 요인이 고려되어야 함. 특히 전화조사 표본에 관한 미국의 최신 동향을 살펴볼 수 있었으며, 이중표본추출 과정의 디자인과 가중치 산출에 관한 유용한 전략을 제시함

2. 일반 세션

가. 모바일 조사의 혁신과 동향

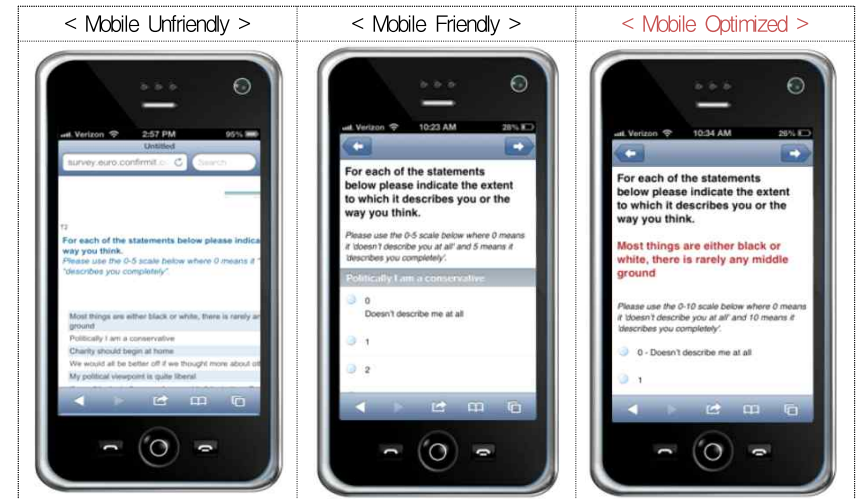
일반 세션

“Mobile Device Survey Research at AAPOR 2015”

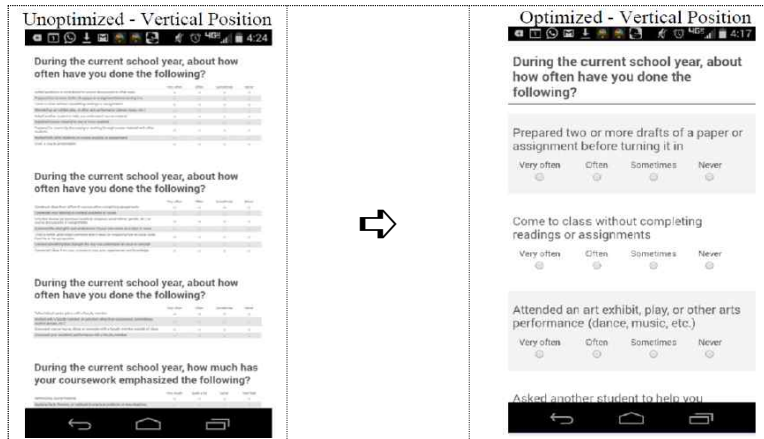
presenter : Benjamin L. Messer, *Research into action portland*

○ **(인터넷과 스마트폰 커버리지)** 자료 수집방법 중 인터넷과 스마트폰 조사의 잠재적인 bias가 있는가? National Survey of Family Growth(‘12.9.-’14.8.) 조사에 의하면 인터넷 가능 82%, 스마트폰 76.1% 수준이며, 교육 및 소득수준이 높은 사람들이 인터넷과 스마트폰 접근성 높음

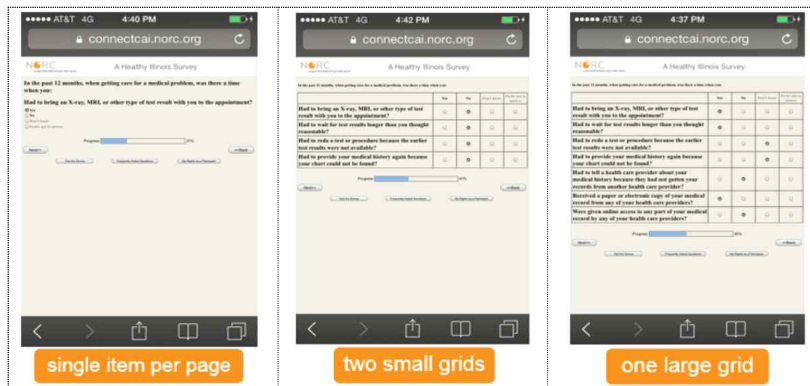
○ **(Mobile에 적합한 디자인)** ‘13년, ‘14년 모바일 조사가 급격하게 각광받게 되면서 모바일 조사 디자인에 관한 연구 수행됨. Mobile Optimized한 디자인은 다른 방법(PC, 태블릿)에 비해 항목이 가장 분리되고, 지속시간이 짧음



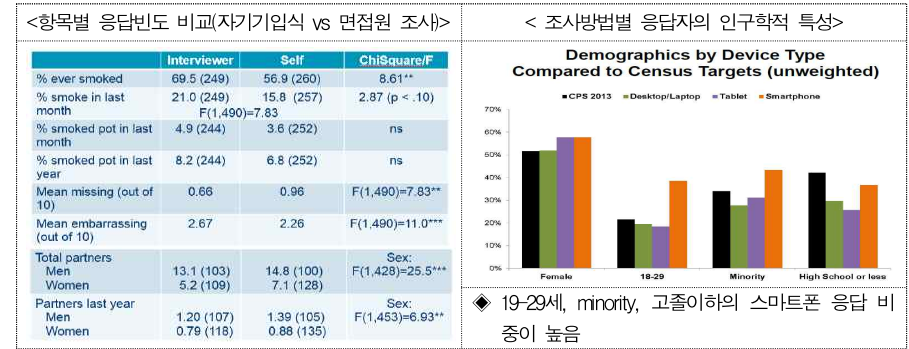
- (긴 조사에서 Mobile 조사에 적합한 디자인) 특히 긴 문항의 설문지는 스마트폰 성격에 맞게 재구성되어야 함. 가독성을 고려하여 항목별로 자르고 풀어서 재설계함* * 2015 National Survey of Student Engagement experiment of ten U.S. colleges/ universities



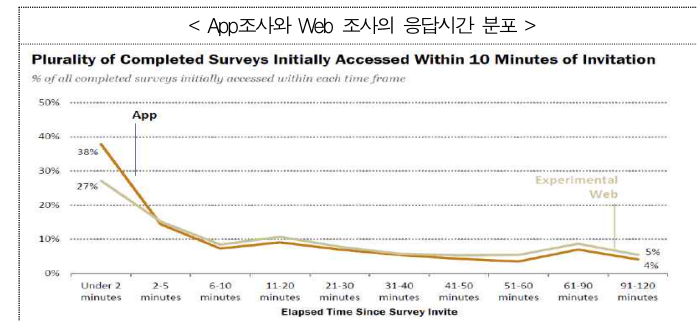
- (모바일 조사에서 Grid의 효과) 자기기입식 조사에서 디자인은 매우 중요하며, 모바일 조사에서 Grid의 효과에 대한 상반된 견해(Pros: 중단 및 응답시간을 감소시킴 Cons: 높은 측정오차, 중단률 증가)가 있음. 3가지 Grid 타입(① single-item/ page, ② two small grids, ③ One large grid)에 대한 실험 결과, Grid 있을 경우 응답시간은 감소했으나, 더 직선적임.(Michael 외, NORC)



- (민감한 항목의 자료수집에 있어서 모바일 조사의 효과) 자기 기입식 조사는 민감한 항목에 있어서 응답이 용이하다는 논의가 있음. 민감한 항목에 대한 자기 기입식 vs 면접원 조사의 결과를 비교한 실험에 의하면, 응답자들은 민감한(당혹스런) 항목에 있어서 스마트폰이나 태블릿의 자기기입식 조사에 더 많이 응답한 것으로 나타남(Roger Tourangeau 외, Westat)



- (모바일 vs 웹조사 간의 응답의 질의 차이) 네덜란드의 social science 패널 종단 인터넷 연구에 의하면, 스마트폰 응답자들이 웹에 비해 정확하게 응답하는 경향이 있으나, 이는 설문내용(민감성, 부담)과 디자인(layout 등)에 크게 영향 받는 것으로 나타남
- (App 조사와 Web 조사의 응답시간 분포) App 조사 응답자가 Web 에 비해 더 빨리 응답하는 경향이 있음. 두 모드 간 응답자의 인구학적 특성 차이는 거의 없으며, 응답하는 장소에 따라 응답에 차이가 발생함



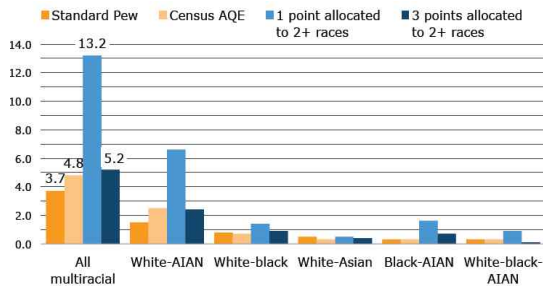
나. 미국 Census 자료에서 Race와 Ethnicity의 측정

일반 세션

“Comparing Methods and Finding of 2020 Census Research on American Indians Responses to Alternative Race Question Instructions”

presenter : Rodney L. Terry, U. S. Census Bureau
e-mail : rodney.terry@census.gov

- (인종의 측정) 미국사회에서 인종구성의 폭발적인 변화는 센서스 조사에서 이들의 정확한 측정에 대한 문제를 야기하였으며(Mixed-race Americans: Who Counts and How Do We Count Them?), 서로 다른 측정에서 다인종 인구 분포는 다르게 나타남



< ACS 조사의 인종 관련 설문 >

Round 1 Version 1: Two questions	Round 1 Version 2: One question
<p>5 Is Person 2 of Hispanic, Latino, or Spanish origin? Mark (X) one or more boxes.</p> <p><input type="checkbox"/> No, not of Hispanic, Latino, or Spanish origin <input type="checkbox"/> Yes, Puerto Rican</p> <p><input type="checkbox"/> Yes, Mexican, Mexican Am., Chicano <input type="checkbox"/> Yes, Cuban</p> <p><input type="checkbox"/> Yes, another Hispanic, Latino, or Spanish origin - Print origin(s), for example, Salvadoran, Dominican, Colombian, and so on. <input type="checkbox"/></p> <p>6 What is Person 2's race? Mark (X) one or more boxes.</p> <p><input type="checkbox"/> White - Print origin(s), for example, German, Lebanese, Egyptian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Black or African Am. - Print origin(s), for example, African American, Jamaican, Nigerian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> American Indian or Alaska Native - Print name of enrolled or principal tribe(s), for example, Navajo Nation, Mayan, Chevak Native Village, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Chinese <input type="checkbox"/> Vietnamese <input type="checkbox"/> Native Hawaiian</p> <p><input type="checkbox"/> Filipino <input type="checkbox"/> Korean <input type="checkbox"/> Samoan</p> <p><input type="checkbox"/> Asian Indian <input type="checkbox"/> Japanese <input type="checkbox"/> Guamanian or Chamorro</p> <p><input type="checkbox"/> Other Asian - Print origin(s), for example, Pakistani, Cambodian, Hmong, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Other Pacific Islander - Print origin(s), for example, Fijian, Marshallese, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Some other race - Print race(s) or origin(s). <input type="checkbox"/></p>	<p>5 What is Person 2's race or origin? Mark (X) one or more boxes AND print the specific race(s) and/or origin(s).</p> <p><input type="checkbox"/> White - Print origin(s), for example, German, Irish, English, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Hispanic, Latino, or Spanish origin - Print origin(s), for example, Mexican or Mexican American, Puerto Rican, Colombian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Black or African Am. - Print origin(s), for example, African American, Jamaican, Nigerian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Asian - Print origin(s), for example, Chinese, Asian Indian, Vietnamese, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> American Indian or Alaska Native - Print name of enrolled or principal tribe(s), for example, Navajo Nation, Mayan, Chevak Native Village, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Middle Eastern or North African - Print origin(s), for example, Lebanese, Iranian, Egyptian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Native Hawaiian or Other Pacific Islander - Print origin(s), for example, Native Hawaiian, Guamanian or Chamorro, Fijian, and so on. <input type="checkbox"/></p> <p><input type="checkbox"/> Some other race or origin - Print race(s) and/or origin(s). <input type="checkbox"/></p>

- 기존버전: 두 개의 문항(히스패닉계 여부/ 인종응답)

- 실험버전: 한 개의 문항

◆ (ACS에서 인종의 측정) monolingual과 bilingual 히스패닉 응답 모두에서 origin와 race의 개념적 불일치 나타남

- (미국 인디언 race와 ethnicity의 측정) 미국 인디언의 인종과 민족의 측정은 매우 복잡하며, 부족의 다양성을 포착하기 힘들. 본 연구는 미국 인디언의 인종과 민족의 측정을 위한 새로운 방법에 관한 심층 면접을 실시한 연구결과를 제시하고 있으며, 후속적이 연구를 통해 2020년 Census를 대비(도입을 목표로 함)

< 미국 인디언 인종과 민족의 측정 >

American Indian Race and Ethnicity Measurement

- Tribal affiliation is currently measured using an open-ended instruction:

American Indian or Alaska Native - Print name of enrolled or principal tribe.
- Current wording is unclear.
- Respondents may be unsure of how to report if their relationship to one or more tribes is complex.

- Open 문항으로 측정
- 워딩이 명료하지 못하며, 두 개 이상의 부족이 섞여 있을 경우 응답하기 힘들
- 워딩은 자기기입식 조사에서 매우 중요한 문제이며, 오픈 문항에서 잘못된 워딩은 폐쇄형 문항보다 자료의 손실을 가져올 수 있음 (Dillman, 2007)

- (Focus group interview 결과) 대부분의 응답자들이 워딩 중 “enrolled or principal tribes(등록된 혹은 주된 부족)”에 불편한 반응을 나타냈으며, 인디언 부족의 복잡한 관계에 의해 복합적인 부족이 보고됨. 실험 버전에서 Write in specific race, origin, or enrolled or principle tribe으로 워딩 수정한 결과, 기존 버전과 측정수준을 달리하는 것으로 인지함

- (요약) 인종에 관한 정확한 자료수집에 워딩의 문제는 매우 중요하며, 후속적인 연구를 통해 설문 문항에 관한 실험버전에 대해 테스트하고, 2020년 센서스에 적용할 예정임

- (시사점) 미국 센서스 국에서는 2020 Census를 대비하여 적응적 조사 설계, 행정 자료의 활용, 무응답 대체 등 관련된 많은 연구가 진행되고 있으며, 체계적인 준비를 하고 있음을 알 수 있었음

다. Beijing Consensus 모델 검증

일반
세션

“Seeking the Beijing Consensus in Asia: An Empirical Test of Soft Power”

Presenter : Jiakun Jack Zhang, UC San Diego

- (연구배경) 중국의 soft power의 팽창에 대한 경험적 검증을 위한 연구로, 워싱턴 컨센서스에 대치되는 베이징 컨센서스(국가 자본주의의 반-민주주의 버전) 중국의 영향력을 공공여론조사를 통해 측정하고자 함

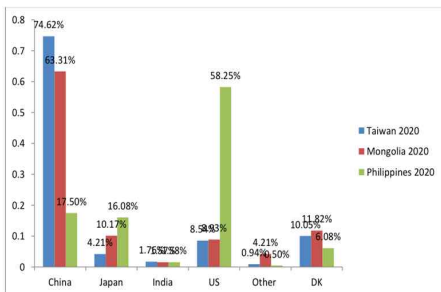
◆ 베이징 컨센서스: 중국 주도의 권위주의 체제 하 시장경제 발전을 일컫는 말로 정치적 자유를 강요하지 않으면서, 시장 경제적 요소를 최대한 도입하는 중국식 발전 모델임. 정부 주도의 점진적 경제개혁과 균형 발전 등의 내용을 강조함

- (활용자료와 변수) 2010년 Asian Barometer Survey*, 독립변인 ① 중국의 주변국에 대한 영향력, ② 중국의 국가 영향력, ③ 중국이 미래 영향력, 종속변인 ① 중국을 발전모델로 선호하는 정도, ② 민주주의 선호의 정도

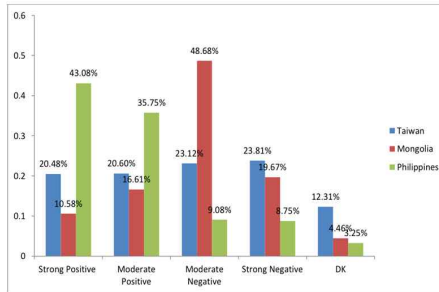
* ABS에서 13개의 동아시아(타이완, 몽골, 필리핀 등) 국가를 대상으로 22개 주제에 관한 서베이(2010년)

- (기초분석) 2020년 아시아에서 가장 영향력있을 국가에 대한 응답 분포는 타이완은 중국 74.6%, 미국 8.5%, 모름 10.1%이며, 몽고는 중국 63.3%, 일본 10.2%, 미국 9.3% 라고 응답했으며, 필리핀은 미국 58.3%, 중국 17.5%, 일본 16.1% 라고 응답함.

< 2020년 아시아에서 가장 영향력있을 국가 >



< 중국 영향력에 대한 태도 >



- (기초분석) 중국의 국가영향력에 대한 동아시아 주변국의 태도는 타이완은 강한 긍정 20.5%, 강한 부정 23.8%이며, 몽고는 강한 긍정 10.6%, 강한 부정 19.7%, 필리핀은 강한 긍정 43.1%, 강한 부정은 8.8% 수준임(몽고의 반중 응답 비중 높음)

- (연구가설) H1a : 중국의 지역 영향력에 긍정적으로 응답한 사람들이 더 중국을 그들의 발전 모델로 생각할 것이다. H1b : 중국의 국가 영향력에 긍정적으로 응답한 사람들이 더 중국을 그들의 발전모델로 생각할 것이다. H1c : 중국의 미래 영향력에 긍정적으로 응답한 사람들이 더 중국을 그들의 발전모델로 생각할 것이다. H2a : 중국의 지역 영향력에 긍정적으로 응답한 사람들이 민주주의에 대한 선호가 낮을 것이다. H2b : 중국의 국가 영향력에 긍정적으로 응답한 사람들이 민주주의에 대한 선호가 낮을 것이다. H2c : 중국의 미래 영향력에 긍정적으로 응답한 사람들이 민주주의에 대한 선호가 낮을 것이다.

< 연구가설 요약 >



- (분석결과) 베이징 컨센서스 이론에 대한 경험적 근거는 약한 수준임. 동아시아 12개 국 응답자들의 중국 영향력에 대한 친밀감은 중국 발전 모델을 채택하는데 약하게 영향을 주는 반면, 중국에 대한 태도와 민주주의 선호와의 관련성은 없는 것으로 나타남. 후속 연구를 통해 경험적 검증을 확장하고자 함(ABS 자료)

< 중국 발전모델 및 민주주의 선호에 대한 로짓 분석 결과 >

Table 1: China vs. Non-China					Table 3: Large Democracy					
EQUATION	VARIABLES	(1)	(2)	(3)	EQUATION	VARIABLES	(1)	(2)	(3)	
		Development Preference	Development Preference	Development Preference			Democracy Preference	Democracy Preference	Democracy Preference	
Development Preference (Per Case Model)	Regional Influence	0.478***			Democracy Preference (4 variable)	Regional Influence	0.0907***			
	Nationalism	0.0453	0.0148	0.125*		Education	0.104***	0.102***	0.0940***	
	Education	0.0642	0.102	0.0620		Income	-0.0289	-0.0278	-0.0271	
	Income	-0.00211	-0.0134	0.00746		Nationalism	0.0698***	0.0686***	0.0807***	
	Country Influence		0.412***			Country Influence		0.0243	0.0233	
	Future Influence			0.0339		Future Influence			0.0129	
	Constant	-1.741***	-1.507***	-3.114***		cut1	Constant	-1.445***	-1.512***	-1.387***
	Observations	3,276	3,545	3,777		cut2	Constant	-0.000476	-0.0633	0.146
	Standard errors in parentheses					Observations	3,087	3,328	3,509	
	*** p<0.01, ** p<0.05, * p<0.1					Standard errors in parentheses				

- (시사점) 베이징 컨센서스에 대한 실증자료를 통한 경험적 검증 시도함. 부상하는 중국 파워의 영향력을 간접적으로 느낄 수 있었음

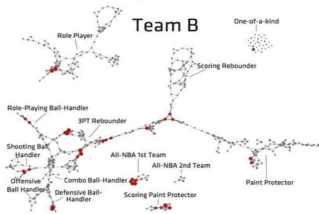
라. Big Data 활용: Text & Sentiment 분석

일반 세션 "Text & Sentiment Analysis of Online Discussion Responses"
 Presenter : Susan Corbelli & Anya Suneson, SMUD

- "Big Data Will Kill Survey Research: Truth or Myth?" 빅 데이터는 조사자료에 비해 커버리지의 낮음, 측정의 소극성, 자료 수집의 낮은 비용 등을 특성으로 함. 조사자료가 기획자의 의도에 의해 기획된 것이라면, 빅 데이터는 비정형화된 자료에서 기획자의 의도에 맞추어 뽑아내는 형태라고 할 수 있음

< 빅 데이터 활용사례와 빅 데이터와 조사자료 간 비교 >

Big Data Analytics in NBA



The above shows 10 positions rather than the customary well established 5 positions.

	Big Data	Surveys
Coverage of total market	Low	High
Passivity of measurement	High	Low
Granularity	High	Low
Transparency	Low	High
Managed, balanced samples	Low	High
Speed of Delivery	High	Low
Availability to all	Low	High
Data collection cost	Low	High

- (소비자 분석의 새로운 형태) 전화, 메일, online survey, Focus group 등의 방법을 전통적으로 사용함. 본 연구에서는 소비자 성향 분석의 새로운 방법론으로 2주 간 SMUD 온라인 게시판에 올려진 자료에 대한 내용 분석을 실시함

< 온라인 게시판 내용 분석 >

Positive Customer Feedback

I appreciate the opportunity to be able to share my opinions and ideas with SMUD. Tons of love for SMUD. That is all =)

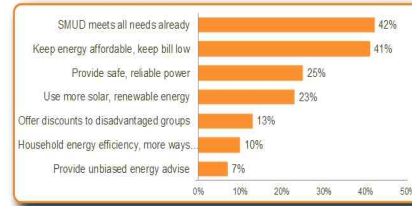
Since I consider SMUD to be a highly ethical company, I would like to be a participant in helping it flourish, in any way I can. I really love that my opinion matters to you.

I appreciate the opportunity to give feedback and voice my concerns, as well as suggestions for future development of my energy provider.

Online Discussion Board Example

- (Text 분석의 방법론) ① Traditional Coding, ② Word clouds, ③ Semantic score (의미 점수), ④ Sentiment Analysis(감성분석)을 활용한 분석사례 제시함

< Traditional Coding >

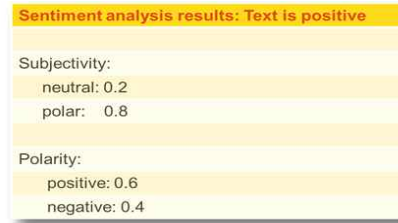


< Word clouds >

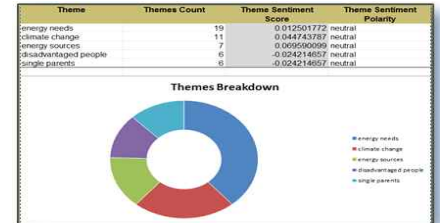


Pros : 빠르고, 쉽고, 시각적 어필함
 Cons: 문장(문맥) 수준에서 분석하지 못함

< Semantic score >



< Sentiment Analysis >



- (감성 분석) 감성분석은 텍스트 마이닝의 한 유형으로 소셜 네트워크 이용자들의 선호에 대한 감성을 분석하는 기법임. 마케팅이나 선거 등에 많이 활용되는 빅 데이터 분석 기술로 단어, 텍스트를 유형화하여 분석함

- (특성) 자료를 신속하게 분석할 수 있으며, 많은 서로 다른 원천의 비정형화된 자료를 활용, 마케팅에 중요한 감성을 측정, 연구자의 주관성이 덜 개입되며, 실시간 분석 가능

- (이점) word cloud에 비해서 복잡적이며, Semantic score와 콤바인 및 문장 수준의 분석 가능함(tool: Semantria).

- (시사점) 공공 여론 측정을 위한 빅 데이터 활용 방안에 관한 많은 이슈와 논의, 특히 마케팅과 선거조사 관련하여 빅 데이터를 많이 활용하고 있으며, 전통적인 조사 방식과의 절충으로 공공여론에 대한 측정에 대한 보다 나은 결과 도출 가능

Ⅲ. 참가소감

- PAPOR는 AAPOR(American Association for Public Opinion Research)의 태평양 지부로 북미 및 캐나다의 여론조사 관련 전문가들이 공공 여론 측정에 관한 연구 성과를 공유하는 장으로서 2012년부터 매년 12월 연례 컨퍼런스를 개최하고 있음
- ‘15년에는 “Measuring Public Opinion in an Expanding Opinion Landscape”를 주제로 하여, RDD 표본설계의 최신 기법, 모바일 조사의 혁신과 동향, 미국 센서스 조사의 설계와 측정, 빅 데이터 활용성 등 공공 여론 측정을 위한 민·관·학 분야 전문가의 다양한 논의가 이뤄짐
- 특히 민간분야에서 활성화된 표본설계 및 혼합조사의 최근 동향, 모바일 조사의 설계와 더불어 다인종 사회에서의 인종과 민족의 측정, 중국의 영향력에 관한 경험적 검증, 시리아 난민 문제 등 인종의 다양성을 담은 공공 여론에 관한 논의가 활성화 됨
- 한편 미국 센서스 국에서는 조사표 설계, 무응답 대체, 행정자료 연계 등 2020년 센서스에 대한 준비를 치밀하고 체계적으로 하고 있으며, 다양한 연구성과를 공유하고 있음을 알 수 있음. 또한 미국 사회에서 공공 여론을 보다 더 잘 측정하기 위하여 민·관·학 합동의 다각도의 노력과 연구 성과의 공유가 매우 자유롭게 활성화되어 있음을 알 수 있었음
- 본 회의에서 입수한 자료는 향후 통계개발원의 분석 및 연구업무에 활용할 예정임(순환센서스 2차 시험조사 혼합조사 도입, 조사항목 재설계)

< 컨퍼런스 전경 >

