2016년 1월 SHPRC 워크샵

10 Years of SHPRC: Survey Sampling and Survey Research

센터장 김 선 웅

Survey & Health Policy Research Center

dongguk UNIVERSITY SHPRC

Outline

1부

동국대 서베이앤헬스폴리시리서치센터의 역사와 사명

2부

Survey Sampling and Survey Research

1. Basic Concepts

2. Popular Sampling Theory

3. Development of Telephone Survey Methodology and Applications

4. Area Probability Sampling using IT

5. Advanced Sampling Methods

1부

동국대 서베이앤헬스폴리시리서치센터의 역사와 사명

Survey Research Center + Health Policy Research Center (Since 2004)

dongguk UNIVERSITY | 서베이앤헬스폴리시리서치센터

Search...

About SHPRC | Center Services | Center Activity | Boards | Links

공지사항 MORE

SHPRC 연구위원회 10월 회의(2013.10.08)	2013.09.18
SHPRC 연구위원회 7월 회의(2013.07.04)	2013.09.18
SHPRC 연구위원회 4월 회의(2013.04.04)	2013.03.28
SHPRC 연구위원회 1월 회의(2013.01.11)	2012.12.18
SHPRC 연구위원회 9월 회의(12.09.21)	2012.10.11

센터소식 MORE

경제활동인구조사 조사방법별 응답자 특성 및 효과적인 CA…	2015.02.11
근로환경조사 조사표 개선 방안 방법론 제안(13.05.02)	2015.02.11
인지면접 사례연구 및 조사표 평가방법(13.04.18)	2015.02.11
지역통계 연구 및 지역별 협력 연구 거점으로서의 역할(12.…	2013.02.27
미국통계학회(2012 JSM) 참가 및 발표(12.07.26~8.4)	2012.10.11

Quick Link

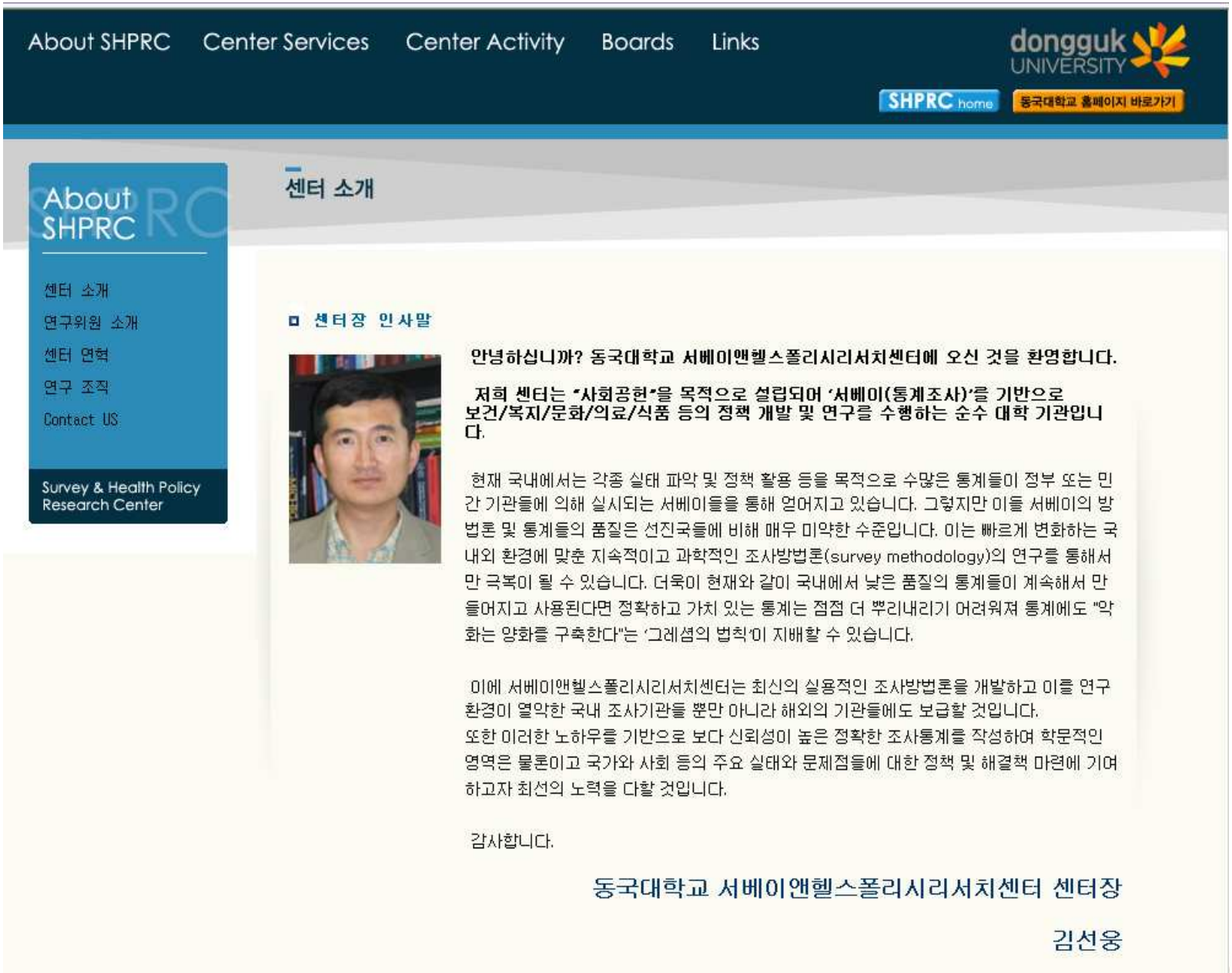
E-CLASS | 사이버강좌 | 중앙도서관

증명서발급 | 유드림스 | 오시는길

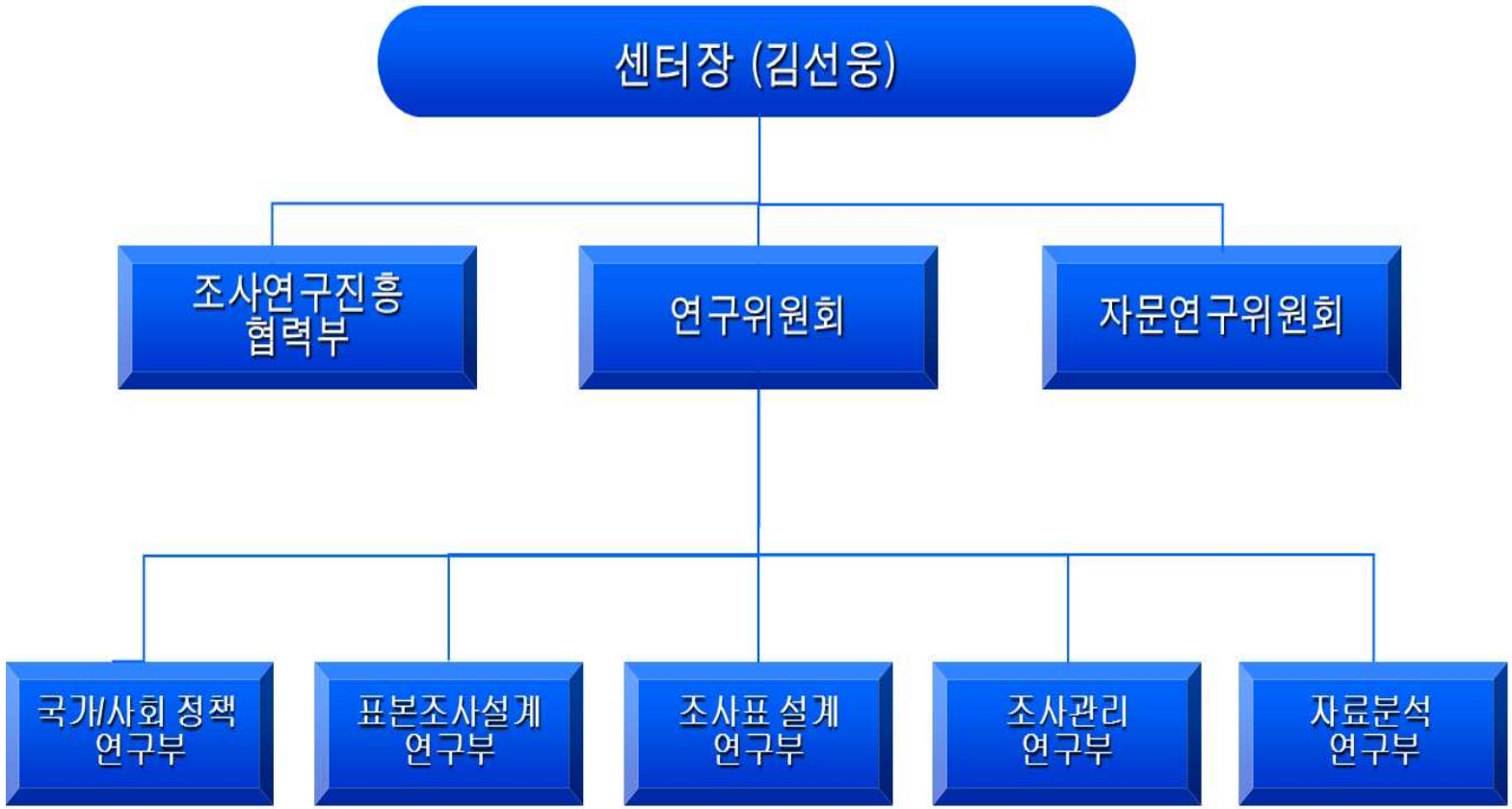
센터 연혁

- 2004. 09. 동국대학교 표본조사연구실(Survey Research Laboratory)
- 2008. 12. 동국대학교 서베이리서치센터(Survey Research Center)로 개칭
- 2011. 10. 동국대학교 서베이앤헬스폴리시리서치센터(Survey & Health Policy Research Center: SHPRC) - 동국대학교 교책연구기관

센터의 사명


About SHPRC
Center Services
Center Activity
Boards
Links
dongguk UNIVERSITY
SHPRC home
동국대학교 홈페이지 바로가기
About SHPRC
센터 소개
연구위원 소개
센터 연혁
연구 조직
Contact US
Survey & Health Policy Research Center
센터 소개
센터장 인사말
안녕하십니까? 동국대학교 서베이앤헬스폴리시리서치센터에 오신 것을 환영합니다.
저희 센터는 "사회공헌"을 목적으로 설립되어 '서베이(통계조사)'를 기반으로 보건/복지/문화/의료/식품 등의 정책 개발 및 연구를 수행하는 순수 대학 기관입니다.
현재 국내에서는 각종 실태 파악 및 정책 활용 등을 목적으로 수많은 통계들이 정부 또는 민간 기관들에 의해 실시되는 서베이들을 통해 얻어지고 있습니다. 그렇지만 이들 서베이의 방법론 및 통계들의 품질은 선진국들에 비해 매우 미약한 수준입니다. 이는 빠르게 변화하는 국내외 환경에 맞춘 지속적이고 과학적인 조사방법론(survey methodology)의 연구를 통해서만 극복이 될 수 있습니다. 더욱이 현재와 같이 국내에서 낮은 품질의 통계들이 계속해서 만들어지고 사용된다면 정확하고 가치 있는 통계는 점점 더 뿌리내리기 어려워져 통계에도 "악화는 양화를 구축한다"는 '그레셤의 법칙'이 지배할 수 있습니다.
이에 서베이앤헬스폴리시리서치센터는 최신의 실용적인 조사방법론을 개발하고 이를 연구환경이 열악한 국내 조사기관들 뿐만 아니라 해외의 기관들에도 보급할 것입니다.
또한 이러한 노하우를 기반으로 보다 신뢰성이 높은 정확한 조사통계를 작성하여 학문적인 영역은 물론이고 국가와 사회 등의 주요 실태와 문제점들에 대한 정책 및 해결책 마련에 기여하고자 최선의 노력을 다할 것입니다.
감사합니다.
동국대학교 서베이앤헬스폴리시리서치센터 센터장
김선웅

동국대학교 서베이앤헬스폴리시리서치센터 조직도



dongguk UNIVERSITY SHPRC

센터 직속 연구위원

유우현 교수(2014 ~)

홍성준 교수(2015 ~)

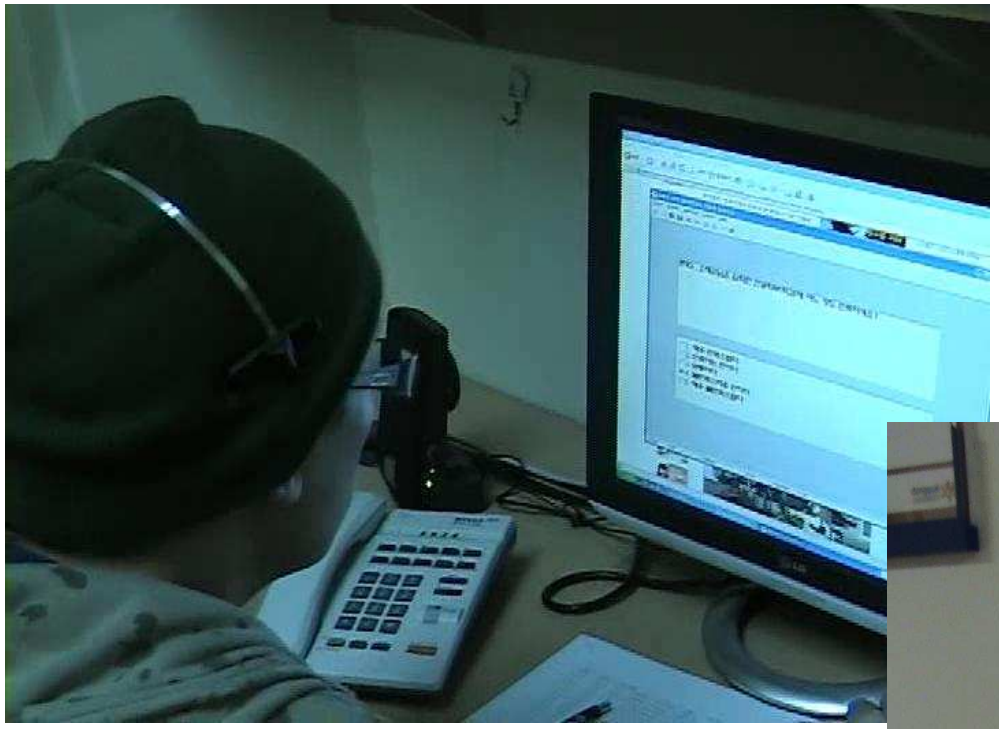
김현태 전문연구위원(2016 ~)

센터 겸직 연구위원

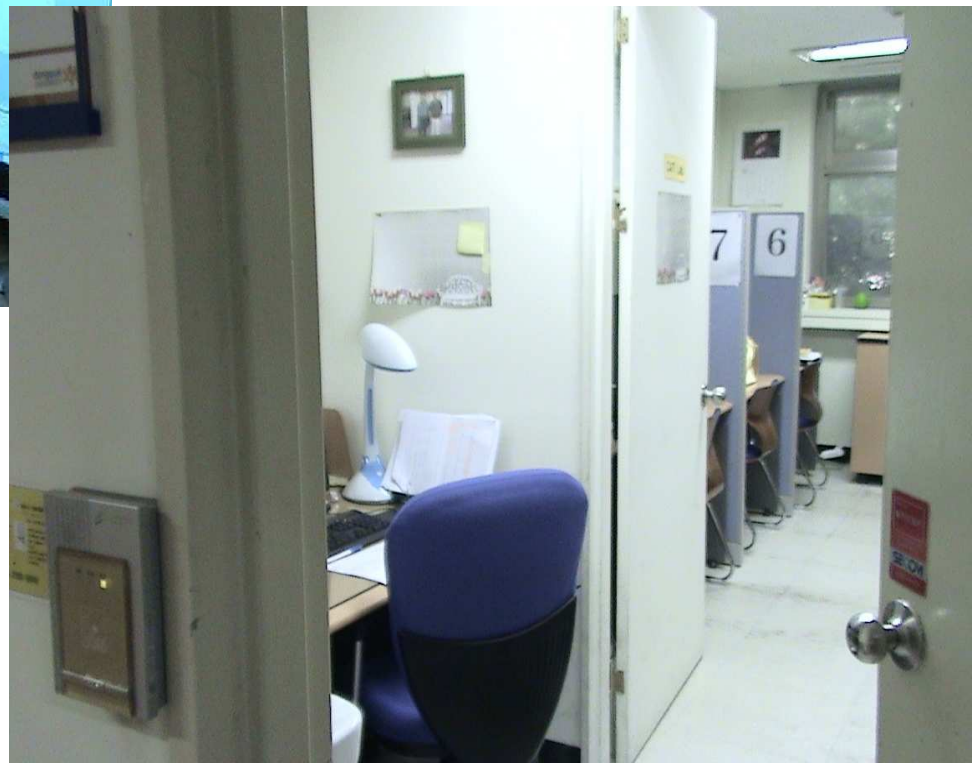
연구위원 소개

	안연순 교수 전공분야 : 직업환경의학 Email : ysahn@dongguk.ac.kr		김학주 교수 전공분야 : 사회복지학 Email : lionking@dongguk.edu
	염용권 교수 전공분야 : 산업공학 Email : yky@dongguk.edu		정경숙 교수 전공분야 : 직업환경의학 Email : bandyoem@naver.com
	염세경 교수 전공분야 : 산업공학 Email : sekyoungyoum@gmail.com		권경희 교수 전공분야 : 사회약학 Email : kkh@dongguk.edu
	김갑순 교수 전공분야 : 회계학 Email : kks@dongguk.edu		정진우 교수 전공분야 : 전자전기공학 Email : kks@dongguk.edu
	김관규 교수 전공분야 : 신문방송학 Email : kwankyu@dongguk.edu		김형용 교수 전공분야 : 사회복지학 Email : fairwork@dongguk.edu

과거 센터 공간(Computer-Assisted Telephone Interviewing)



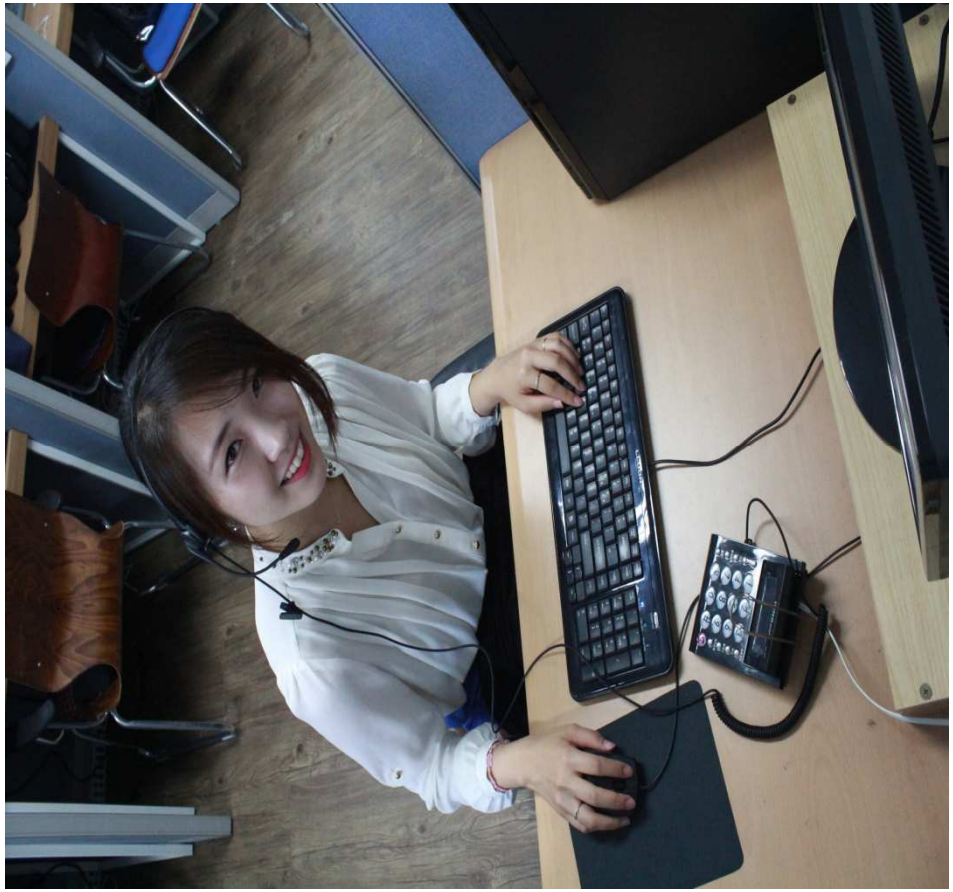
2004년 ~ 2007년 (2.5평)

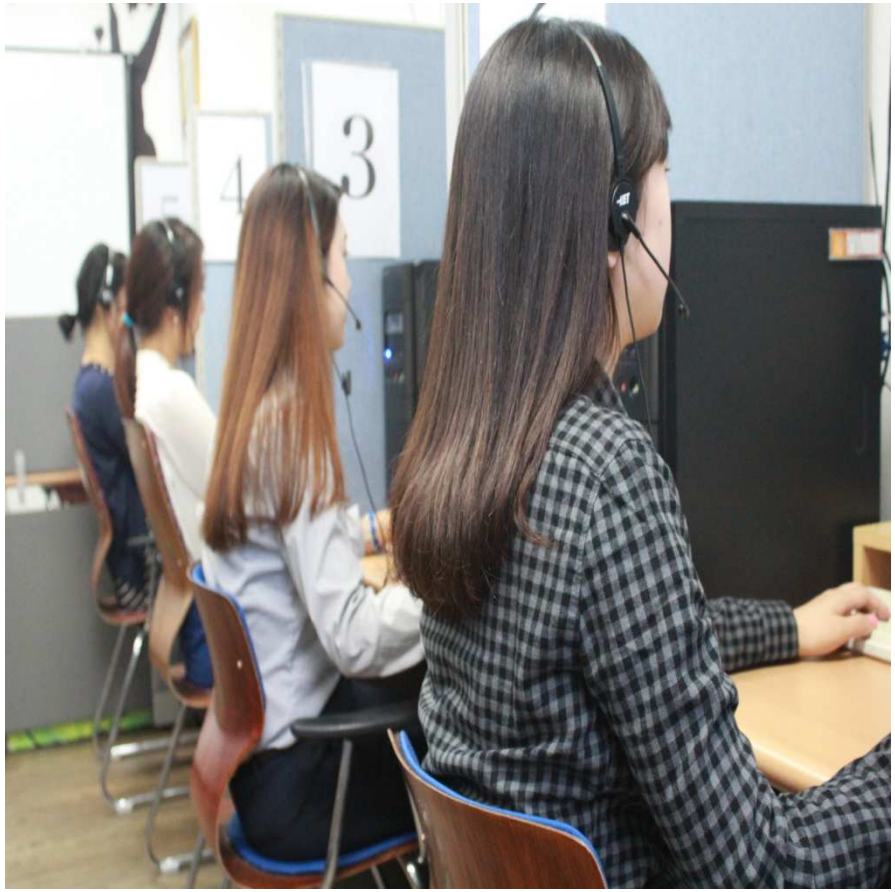


2008년 ~ 2011년(5평)

dongguk UNIVERSITY SHPRC

현재 센터 공간(Computer-Assisted Telephone Interviewing)





2017년~ 센터 공간?

dongguk UNIVERSITY SHPRC

2004 ~ 2015

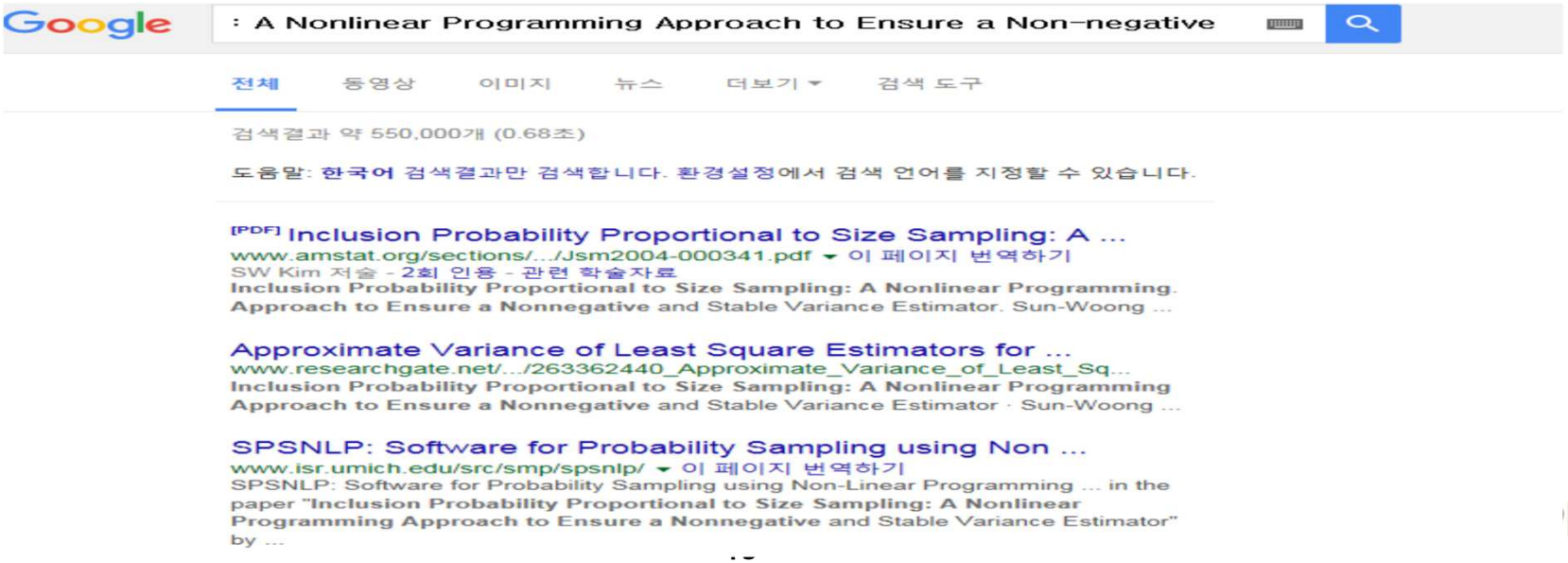
조사연구 프로젝트
조사방법론 연구논문
조사연구결과 논문

2004 Research Projects and Selected Papers

한·양방 복합투약에 관한 조사연구, 보건복지부

이동전화 확산에 따른 유선전화 가구보유율의 변화: 한국을 포함한 주요국가들을 중심으로, 조사연구, 한국조사연구학회

Inclusion Probability Proportional to Size Sampling: A Nonlinear Programming Approach to Ensure a Non-negative and Stable Variance Estimator, *Joint Statistical Meetings (American Statistical Association)*, Toronto, Canada



2005

퇴행성 슬관절염의 맞춤형 침구 임상시험 프로토콜 개발을 위한 전화조사, 대한침구학회지

퇴행성 슬관절염의 맞춤형 침구 임상시험 프로토콜 개발을 위한 면접조사, 대한침구학회지

Dual-frame Landline/Cellular Telephone Survey Design, *Joint Statistical Meetings*, Minneapolis – 세계 최초 일반전화/핸드폰조사 설계 연구
– University of Michigan(ISR)과 공동연구

2006

국내 자락(사혈)요법 임상실태 조사연구, 한국한의학연구원

Model-based Sampling Designs for Optimum Estimation, *Joint Statistical Meetings*, Seattle

2007

국내 자락(사혈)요법 임상 실태 파악을 위한 면접조사, 대한침구학회지

Sample Allocation under a Population Model and Stratified Inclusion Probability Proportional to Size, *Joint Statistical Meetings*, Salt Lake City

2008

Encyclopedia of Survey Research Methods, Sage Publications

National Telephone Survey: Implementing List Assisted RDD Design and Randomizing Respondent Selection Process, 한국조사연구학회

A Comparison for Data Collection Modes: Computer-Assisted Telephone Interviewing and Paper-and-Pencil Self-Interviewing, 한국조사연구학회

Optimized Whole Sample Procedures vs. Traditional Draw-by-Draw Procedures, *Joint Statistical Meetings*, Denver

2008 임금제도 실태조사 표본설계, 한국노동연구원

2009

레이져 뜸의 수요예측을 위한 시장조사, 한국보건산업진흥원

Some Methods of Model-Based Sampling, Joint Statistical Meetings

List-Assisted RDD Sampling in Korea: Testing the Feasibility of National Surveys, *World Association for Public Opinion Research*, Lausanne, Switzerland

2010

RDD 표본추출 기법 기술이전 사업, 동국대 - 현대리서치연구소

국내 뜸요법 사용실태에 관한 표본조사 결과, 대한침구학회지

2011

지역사회에서 나트륨 섭취량 조사와 나트륨 섭취량이 혈압과 동맥경화에 미치는 영향 평가, 식품의약품안전청

Patients' Assessment of Community Primary and Non-primary Care Physicians in Seoul City of South Korea, 가정의학회지

한의학 연구에서 표본조사를 위한 가이드라인, 대한한의학회지

Using Telephone Surveys in Medical Researches, 대한가정의학회

Easy or Hard Choices?: Faulty RDD, Near RDD and True RDD Telephone Sampling, 한국조사연구학회

Why Did the Pre-Election Polls in 2010 Local Elections Go All Wrong?, 한국조사연구학회

Comparing Cell Phone and Web for a Student Survey, *Joint Statistical Meetings*, Miami Beach, Florida

2012

일반 산단 주변지역 환경보건평가, 국립환경과학원

취업자 근로환경조사 심층분석 연구, 산업안전보건연구원

1. List-Assisted RDD Sampling in Korea: Testing the Feasibility of National Survey Under Within-Household Selection, *International Journal of Public Opinion Research* 2012

2. A Simple Approach to Sample Allocation for Multivariate Stratified Sampling, presented in the *Joint Statistical Meetings*, San Diego, California 2012

3. Using New IT in Area Sampling: An Experience in Korea, presented in the *Joint Statistical Meetings*, San Diego, California 2012

4. Estimation of Salt Intake using List-Assisted RDD, presented in the *Joint Statistical Meetings*, San Diego, California 2012

5. New Model-Optimized Sampling Techniques, presented in the *Joint Statistical Meetings*, San Diego, California 2012

6. A Comparison of Response Patterns between Landline and Cell Phone RDD Surveys, *World Association for Public Opinion Research*, 2012

7. Constructing Hard to Survey Index in the Korean Labor Force Survey, *World Association for Public Opinion Research*, 2012

8. Using Cell Phone for Student Surveys, *World Association for Public Opinion Research*, 2012

9. 2010-2012년 국내 선거여론조사에서의 오차, 한국조사연구학회 춘계학술대회, 2012

10. Optimizing Call Scheduling in a Dual Frame RDD Survey, 한국조사연구학회 추계학술대회, 2012

2013

Telephone Survey of Adults during the S. Korean Presidential Campaing, University of Michigan

근로환경조사의 정확성 제고를 위한 조사방법론 연구, 산업안전보건연구원

Sodium Intake of a City Population in Korea Estimated by 20-h Urine Collection Method, *European Journal of Clinical Nutrition*

Using New IT for Area Sampling in a Metropolitan Household Survey, *Joint Statistical Meetings*, Montréal, Canada

Reducing Survey Nonresponse through Enhanced Administrative Cooperation: an Experience in Korea, *Joint Statistical Meetings*, Montréal, Canada

2012 대통령선거 Two-Wave Dual Frame RDD 전화 패널 조사, 한국조사연구학회

Estimating Prevalence, Awareness, and Treatment of Hypertension using List-Assisted RDD Methodology 한국통계학회

2014

2014 마약류중독자 실태조사, 보건복지부 – 국내 최초의 일반가구 대상 조사

국내 핸드폰 조사의 현재와 미래: RDD 표본의 효율성과 응답자 분포를 중심으로, 한국조사연구학회

조사표 평가방법의 효율성: 전문가 검토와 인지면접의 비교, 한국조사연구학회

Evaluating Coverage and Accuracy of 2010 Korea Census List of Households using New IT, 한국조사연구학회

Using Two-Wave Dual Frame RDD Telephone Pre-election Poll in the 2012 Korean Presidential Election, *World Association for Public Opinion Research*, Nice, France – 세계최초 일반전화/핸드폰 패널조사 연구

2015

2015 환경소음공해 건강영향 실태조사, 한국환경산업기술원

2015 지역사회건강조사 조사 모니터링 개선 방안 연구, 질병관리본부

식품류 곰팡이독소 오염평가방법 연구(1차년도), 식품의약품안전처

마약류중독자 실태조사와 정신질환 역학실태조사 연계방안 마련 연구, 보건복지부

6.4 지방 선거 이후, 여론 조사 무엇이 달라질 수 있나?: 선거 여론 조사 등록 자료를 중심으로, 한국조사연구학회

Reducing Unit Nonresponse in Controlled Access Situations: An Experimental Study in South Korea, *Joint Statistical Meetings*, Seattle

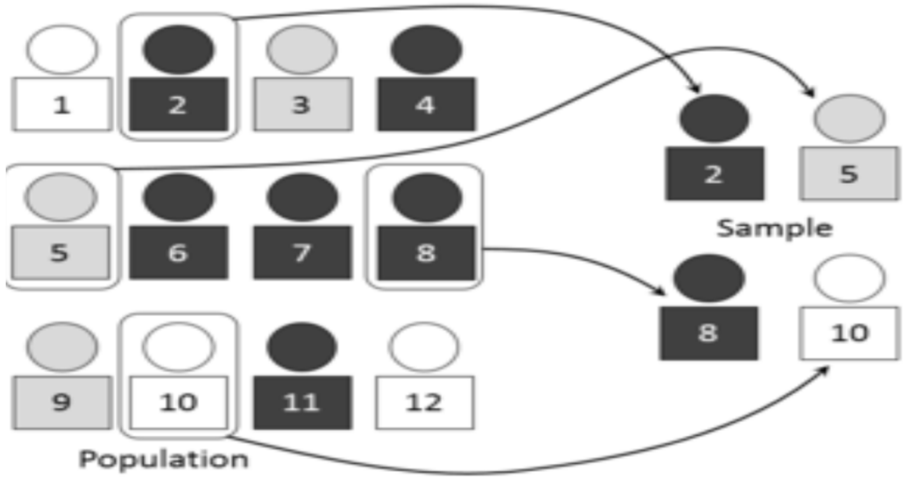
2부

Survey Sampling and Survey Resarch

1. Basic Concepts

Sampling

The process of selecting a subset of population



Survey Sampling

The process of selecting a random sample of elements from a ***frame*** to conduct a ***survey*** for a ***target population***

What do mean “random”?

It means that all human influence is removed from the selection process.
To achieve a random selection, random numbers are used.

The samples corresponding to those random numbers selected are called “probability samples.”

Probability or Non-probability Sample?

1) “Survey 2000” by the US National Geographic Society (NGS)

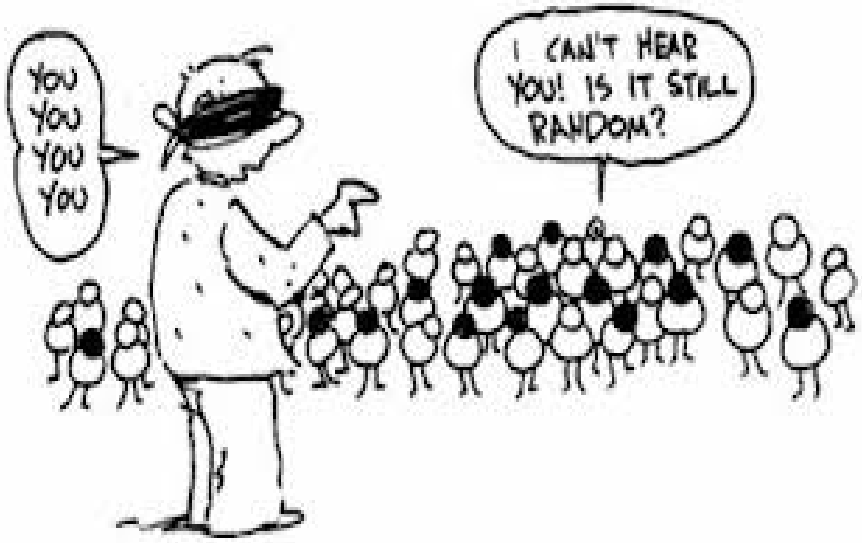
- Web-based survey
- 80,000 visitors to the web page and over 50,000 completions of the questionnaire
- Seeing live theater: 60%
- Visiting an art museum or gallery: 77%

2) “Survey of Public Participation in the Arts” by the US National Endowment for the Arts

- Telephone survey
- Using randomly generated telephone numbers
- Musical theater: 25%
- Nonmusical theater: 16%
- Visiting an art museum or gallery: 35%

Is Sampling Important in Survey?

If the sample is selected haphazardly or with subjective decision, there may be little hope of making correct estimation to the population, although other survey processes including field supervision and management are successfully completed.



The haphazard or convenience selection methods have a disadvantage – there is no theoretical support. Many surveys still use haphazardly or purposively selected samples.

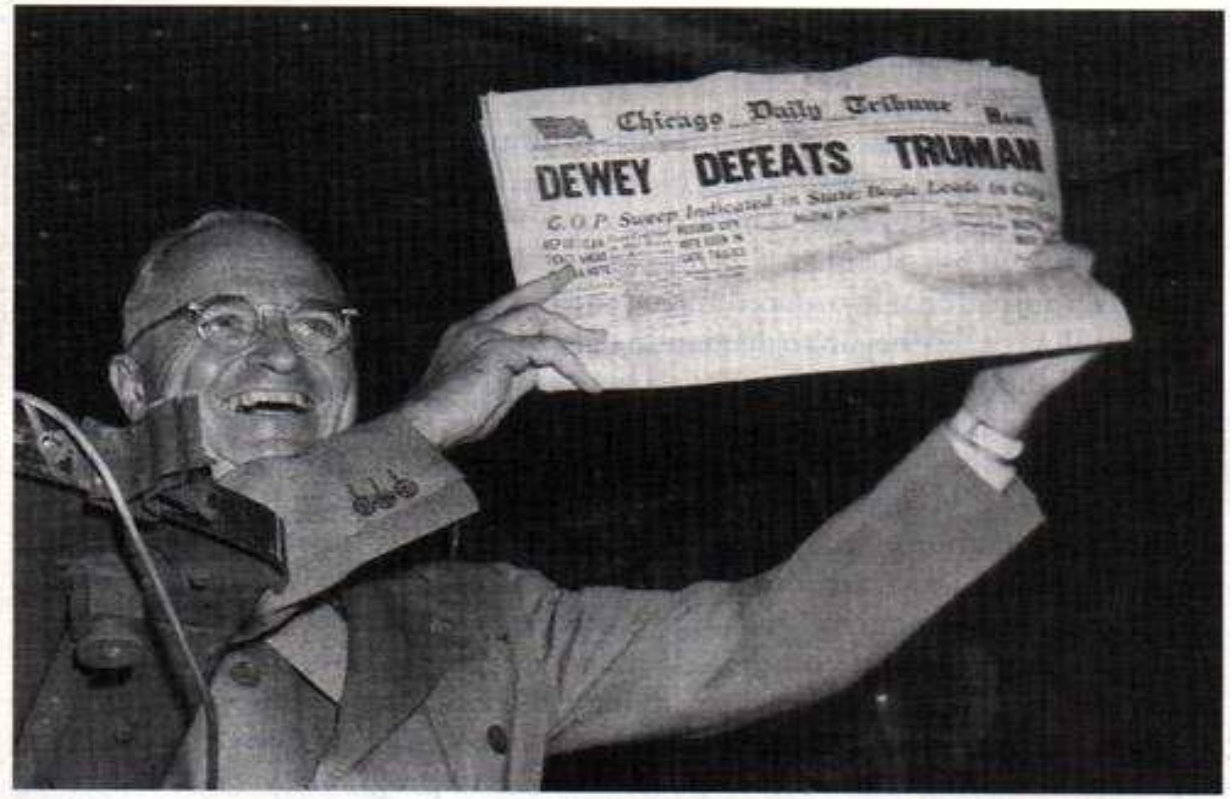
Is Sampling Important in Survey? (Cont.)

(Example) Quota Sampling

Control Characteristic	Population composition	Sample composition	
	Percentage	Percentage	Number
Sex			
Male	48	48	480
Female	52	52	520
	100	100	1000

Is Sampling Important in Survey? (Cont.)

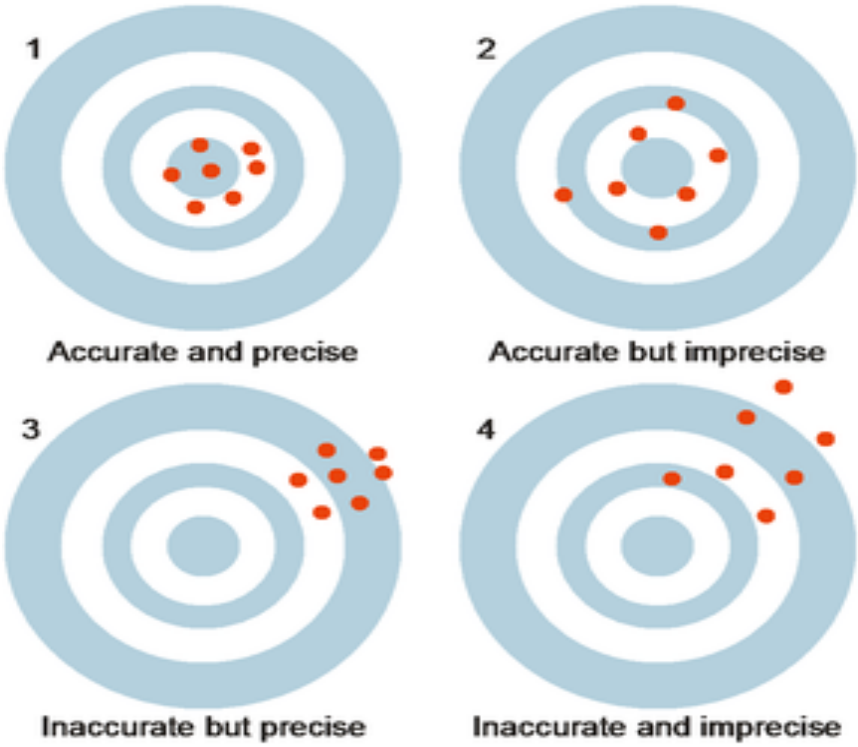
U.S. Presidential Election 1948



Correct Estimation?

1) Accuracy: sampling bias
2) Precision: sampling variance

They arise through sampling and can be controlled by sampling.



Samples and Estimates

If one use a single probability sample, inferences about the population with known levels of confidence can be made

A confidence interval is commonly denoted by

$$\overline{y} \pm z_{\alpha/2} se(\overline{y})$$

$$p \pm z_{\alpha/2} se(p)$$

Survey Statistics and Survey Statisticians

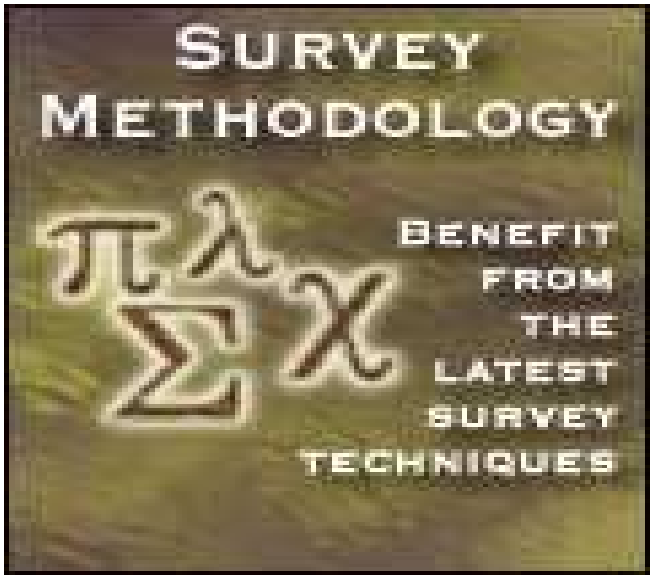
Survey Statistics (조사통계학)

Survey Statisticians (조사통계학자) or Survey Methodologists (조사방법론학자)

Sample Designs (표본설계) or Sampling

Samplers (표본설계전문가)

AAPOR and the American Statistical Association



Statistics Canada

Jerzy Neyman?

- Randomized Experiments in 1923
- Confidence Interval in 1937
- Neyman-Pearson Lemma for hypothesis testing
- The groundbreaking event leading to modern scientific sampling

Neyman, J. (1934), " On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, 97, pp. 558-625.

Components of Sample Design

1) ***Sample size***

2) ***Selection probabilities*** of elements or samples

3) ***Stratification*** to control the representation of key subpopulations in the sample

4) Selection of groups called ***clusters***

2. Popular Sampling Theory

Simple Random Sampling

This sample design assign an ***equal probability of selection*** to each frame element ($n=1$), to all pairs of frame elements ($n=2$), to all triplets of frame elements ($n=3$), etc.

(Procedure)

1) Number all elements within a frame from 1 to N.
2) Select random numbers from 1 to N without replacement.
3) Identify the elements corresponding to the selected random numbers.

$$p(s)=\frac{n}{N}\frac{n-1}{N-1}\frac{n-2}{N-2}\cdots\frac{1}{N-n+1}=\frac{n!(N-n)!}{N!}=\frac{1}{{}_{N}C_{n}}$$

Simple Random Sampling (Cont.)

SRS is a special case of an “epsem” for ***e***qual ***p***robability of ***se***lection **m**ethod.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$v\left(\bar{y}\right) = (1-f)\frac{s^2}{n}, \quad f = \frac{n}{N}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

(Theorem)
The sample mean $\bar{y}$ from a simple random sample is an unbiased estimator of μ

(pf)
$$E(\bar{y}) = \sum_s \bar{y} p(s)$$
$$= \sum \bar{y} \frac{1}{{}_N C_n}$$
$$= \frac{n!(N-n)!}{N!} \frac{1}{n} \sum (y_1 + y_2 + \cdots + y_n)$$
$$= \frac{(n-1)!(N-n)!}{N!} \sum (y_1 + y_2 + \cdots + y_n)$$
$$= \frac{(n-1)!(N-n)!}{N!} {}_{N-1}C_{n-1} (y_1 + y_2 + \cdots + y_N)$$
$$= \frac{(n-1)!(N-n)!}{N!} \frac{(N-1)!}{(n-1)!(N-n)!} (y_1 + y_2 + \cdots + y_N)$$
$$= \frac{1}{N} (y_1 + y_2 + \cdots + y_N)$$

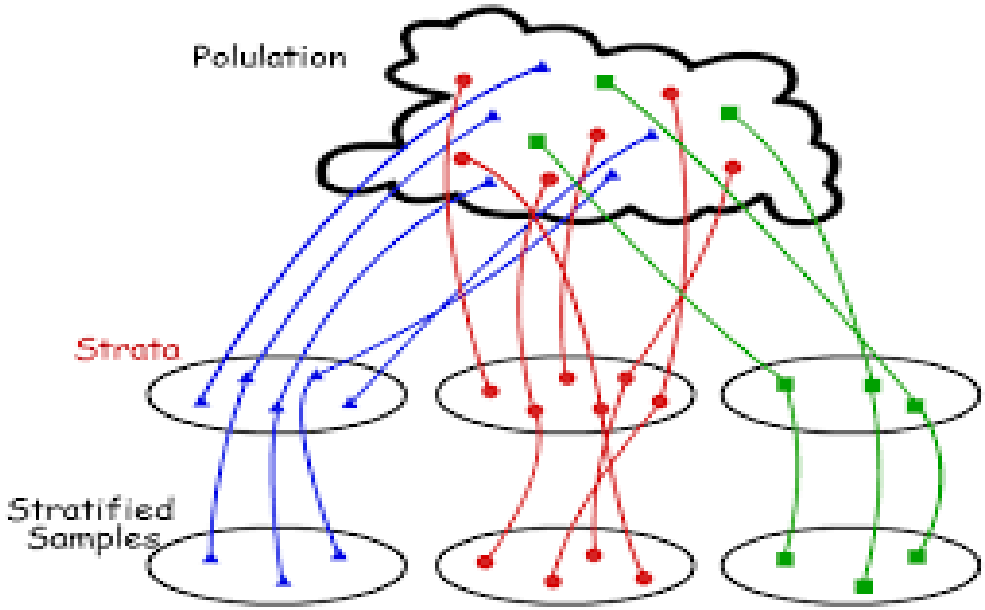
dongguk UNIVERSITY SHPRC

Stratified Random Sampling

Each element in a population (a frame) can be placed in a single “stratum (group).”

Separate samples are drawn from each stratum, using the same sampling method or using different sampling methods.

Separate estimates are obtained from each stratum, and they are combined to estimate population values.



Stratified Random Sampling (Cont.)

Sampling fraction in epsem: $f = \frac{n}{N}$

Proportionate allocation to select the sample in each stratum h with SRS:

$$f_h = \frac{n_h}{N_h} = f \Rightarrow n_h = n\frac{N_h}{N} = nW_h \Rightarrow n_h \propto N_h$$

(Example) From $N = 8,000$ schools divided into three strata, a stratified sample of $n = 480$ is selected. The survey asks principals of the schools how many vocational classes are offered this year.

$$n_h = 480\frac{N_h}{8,000} = 0.06N_h$$

Unweighted mean: $(6+5+8)/3 = 6.3$

Weighted mean: $\bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h$

$$= (0.4\times 6) + (0.5\times 5) + (0.1\times 8)$$

$$= 5.7$$

Sampling variance: $v\left(\bar{y}_{st}\right) = \sum_{h=1}^{H} W_h^2 (1-f_h)\frac{s_h^2}{n_h}$

$$s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2$$

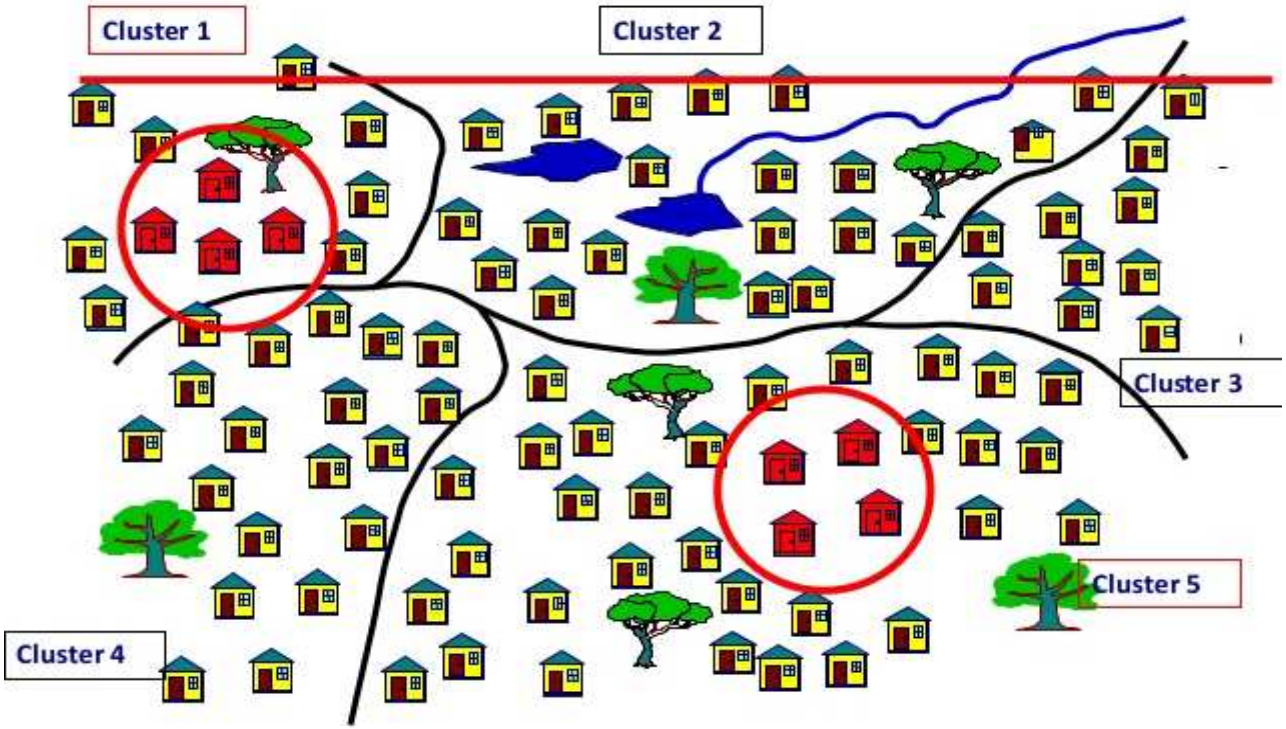
$$se(\bar{y}_{st}) = \sqrt{v(\bar{y}_{st})} = \sqrt{0.00920} = 0.096$$

95% Confidence interval estimate: $5.7 \pm 1.96 \cdot 0.096$ or (5.5, 5.9)

dongguk UNIVERSITY SHPRC

Cluster Sampling

We may sample clusters of elements and then do interview the elements only within the selected clusters.



dongguk UNIVERSITY SHPRC

$$\bar{y} = \frac{1}{aB}\sum_{\alpha=1}^{a}\sum_{\beta=1}^{B} y_{\alpha\beta}$$

$$v(\bar{y}) = (1-f)\frac{s_a^2}{a}$$

$$s_a^2 = \frac{1}{a-1}\sum_{\alpha=1}^{a}\left(\bar{y}_\alpha - \bar{y}\right)^2$$

3. Development of Telephone Survey Methodology and Applications

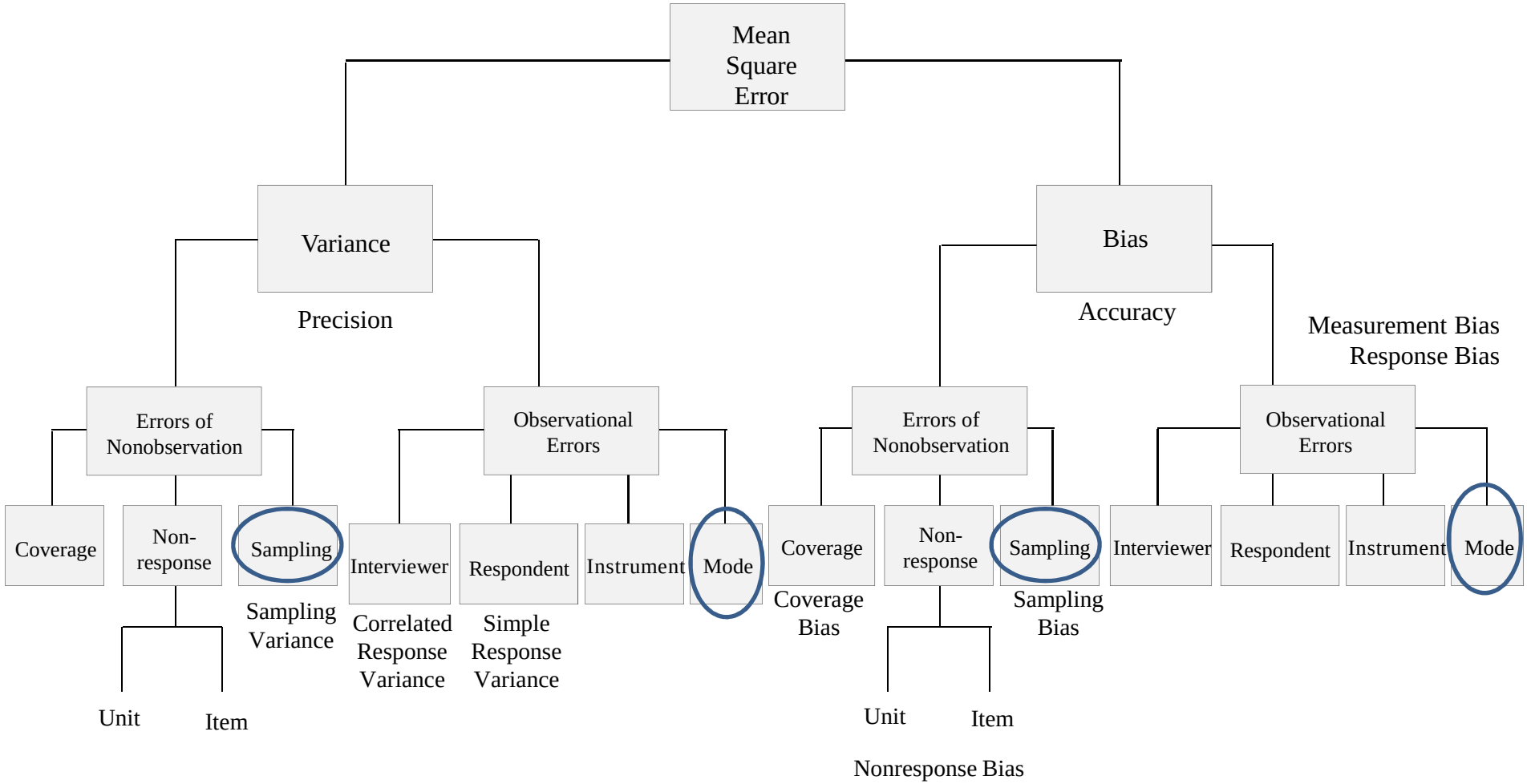
Groves *et al*. (2009)

> "***Survey methodology*** seeks to identify principles about the ***design***, ***collection***, ***processing***, and ***analysis*** of surveys that are linked to the cost and quality of survey estimates."

Information in a survey is gathered primarily by asking people questions.

Sources of Errors in Surveys

❖ The structure and language of errors in total survey error (Groves, 2004)



dongguk UNIVERSITY SHPRC

Face-to-Face (FTF) Surveys and Area Frames

Changing Mode (Cont.)

By the late 1960s, the cost of face-to-face (FTF) surveys was escalating while, at the same time, the proportion of households with telephones had grown to close to 90 percent, both in North America and Europe.

Furthermore, the decline in response rates in FTF surveys made the possibility of using the telephone as a mode.