

# 표본론 및 실습

: Lec.1 개요(Introduction)

- 이 인 규 -



# Key Point

---

- ◆ 자료수집 방법 : 조사 vs 실험
- ◆ 조사의 기본 절차
- ◆ 목표모집단과 조사모집단
- ◆ 오차의 종류

(ref.)

- \* (박민규, 강현철) chap.1
- \* (김재광) chap.1

# 통계학(statistics)이란...

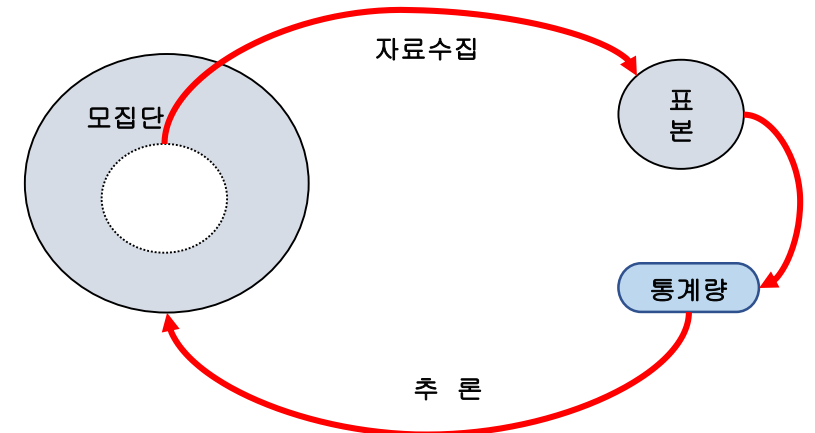
## ◆ (정의) 통계를 잘 다루기 위한 학문

사회, 자연현상을  
계량화한 정보

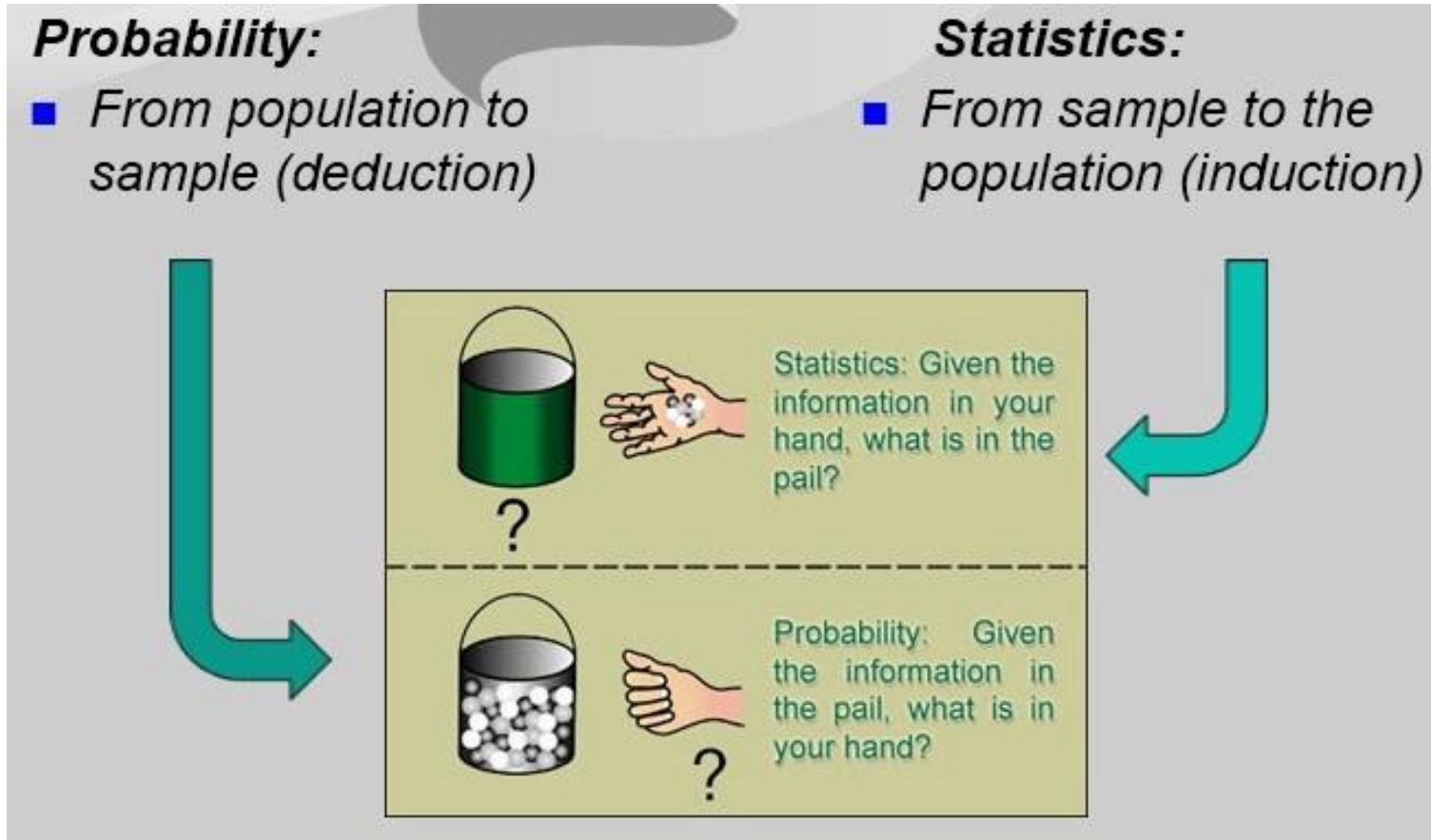
(자료 수집) 자료의 수집과정을 설계하고,  
(정리·요약) 자료를 정리 및 요약하여,  
(분석·해석) 추정, 가설검정 등 자료를 분석 및 해석

## ◆ (목표) 합리적이고 객관적인 자료수집과 분석방법 제공

## ◆ (목적) 자료에 내재된 유의미한 정보를 도출하여, 세상에 대한 이해를 넓혀가는 데 기여



# 통계학과 수학(확률론)의 차이



Source : Ranjith Kumar Ramasamy  
(<https://www.linkedin.com/pulse/sampling-distribution-models-central-limit-theorem-ranjith-kumar>)

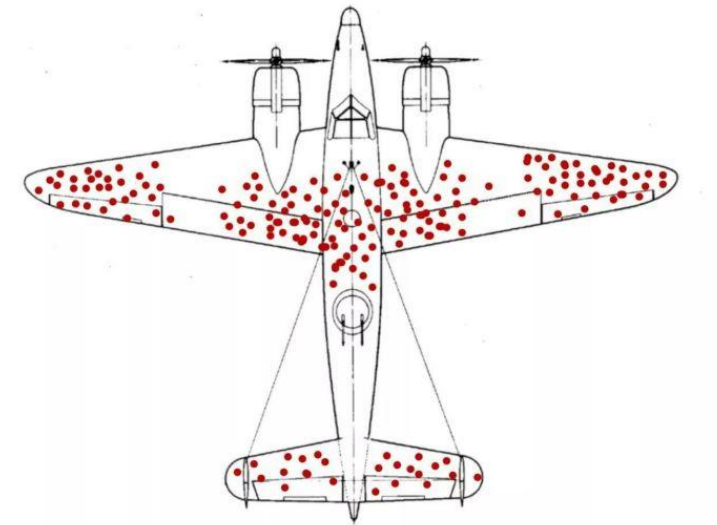
# 자료가 갖추어야 할 요건

## ◆ 신뢰할 수 있는 자료란...

- \* (목적에 부합) 분석하고자 하는 목적에 합당해야 하고
- \* (객관성) 자료수집 과정에 주관적 편견이 배제되어야 하며
- \* (대표성) 연구 대상 집단을 잘 대표해야 한다.

## ◆ 선택편향(selection bias)

- \* 목표집단(target)과 조사집단(sample)의 불일치
- \* 행정자료와 같은 보고자료(recording data) 역시  
선택편향 또는 대표성(representativeness)에 한계가 있음



(그림 1) 2차 세계대전 당시 생존 전투기의 탄흔 기록

# 자료수집 방법 : 조사 vs 실험

---

## ◆ 조사(survey)

- \* 주어진 현상에 대한 관찰(observation)
- \* 현상이나 사건 사이의 상관관계 탐색
- \* 조사 대상에 대한 인위적 조절이 불가능하며 통상 대상 집단의 총 개체가 유한(finite)함

## ◆ 실험(experiment)

- \* 실험의 대상, 즉 처리 결과의 측정(measurement)
- \* 요인과 반응 사이의 인과관계 규명
- \* 실험 대상에 대한 인위적 조절이 가능하며 대체로 조절 가능한 수준의 수가 무한(infinite)함

⇒ 조사와 실험의 차이 : 인과성 여부

“흡연자 9명당 1명꼴로 폐암이 관측되었다” 흡연과 폐암의 인과성을 단정할 수 있을까?

# 자료수집 방법 : How to get the data?

	조사 연구 (survey study)	실험 연구 (experimental study)
주요 목적	현상 및 상관관계 탐색 (description or correlation)	인과관계의 검증 (causality)
통제 가능성	낮음	높음
활용분야	사회과학, 마케팅, 여론조사 등	심리학, 의학, 교육학 등
예시	두 종류 광고의 효과	두 종류 약물의 효과
key point	어떻게 효과적인 표본을 추출할 것인가? (sampling design)	어떻게 외부요인의 영향을 통제할 것인가? (experimental design)

⇒ 연구의 목적과 주어진 조건에 합당한 자료 수집 방안의 설계가 필요

# 조선 최초의 여론조사

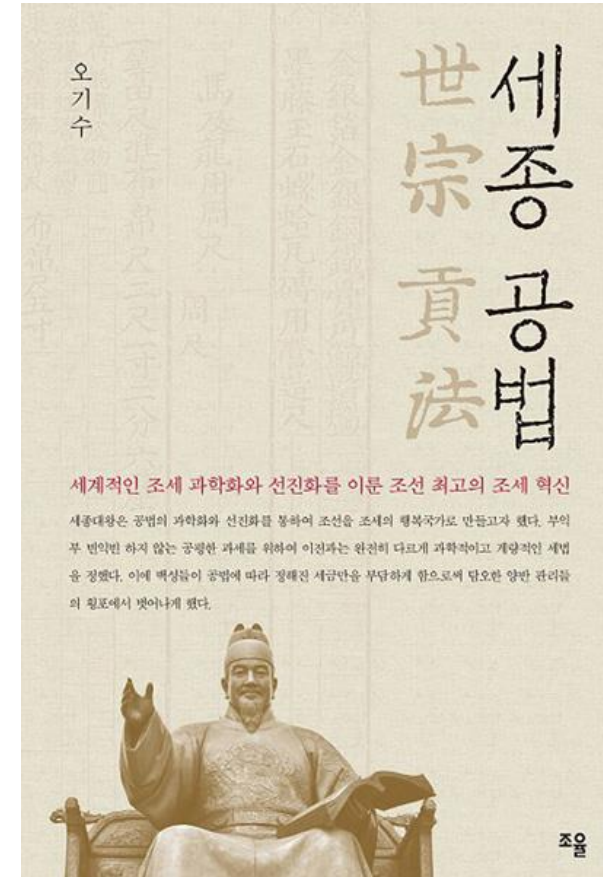
## ◆ '공법' 시행에 대한 전 백성 여론조사(세종 12년, 1430년)

- 표본크기 : 172,806명(전국민의 1/4)
- 조사기간 : 5개월

지역	찬성(명)			반대(명)		
	수령	백성	%	수령	백성	%
경기도	29	17076	99	5	236	1
유후사(개성)		1123	94.1		71	5.9
평안도	6	1326	4.5	36	28474	95.5
황해도	17	4454	22.3	17	15601	77.7
충청도	13	6982	33	26	14013	67
강원도	5	939	12.6	10	6888	87.4
함길도	3	75	1	14	7387	99
경상도	55	36262	99	16	377	1
전라도	42	29505	99	12	257	1
합계	170	97742		135	73264	
	97912명			73384명		
	(전현직 관료 관찰사 및 그 외포함 총 98657명)			(전현직 관료 관찰사 및 그 외 포함 총 74149명)		

<세종실록> 1430년 8월10일자에 실린 공법 관련 기사에서 여론조사를 분석한 각 도별 결과. 전라와 경상 등 상대적으로 전답이 많고 비옥한 지역의 찬성률이 높았고, 강원 평안 함길 등 척박한 지역일수록 반대여론이 높았다. |소진령의 논문에서 인용한 표를 재정리한 것

(경향신문 : 이기환의 흔적의 역사)



\* 공법 : 토지의 비옥도에 따라 등급을 나누고, 매년의 풍흉에 따라 세금 기준을 정해 일정한 세금을 부과



# 조사의 기본 절차(Survey Process)

① 조사 목적 및 연구 대상 정의

② 조사 설계(project design)

③ 자료 수집(data collection)

④ 자료 처리(data cording)

⑤ 조사 분석(analysis)

⑥ 보고(reporting)

- 조사의 목적(필요성) 및 연구의 범위 명확화
- **목표모집단(target population)**의 구체화  
: 조사 및 추론의 결과를 직접 적용할 수 있는 대상 집단
- 사용된 용어의 정확한 의미  
(ex) "결손가정 청소년의 영양 실태"

# 조사의 기본 절차(Survey Process)

① 조사 목적 및 연구 대상 정의

② 조사 설계(project design)

③ 자료 수집(data collection)

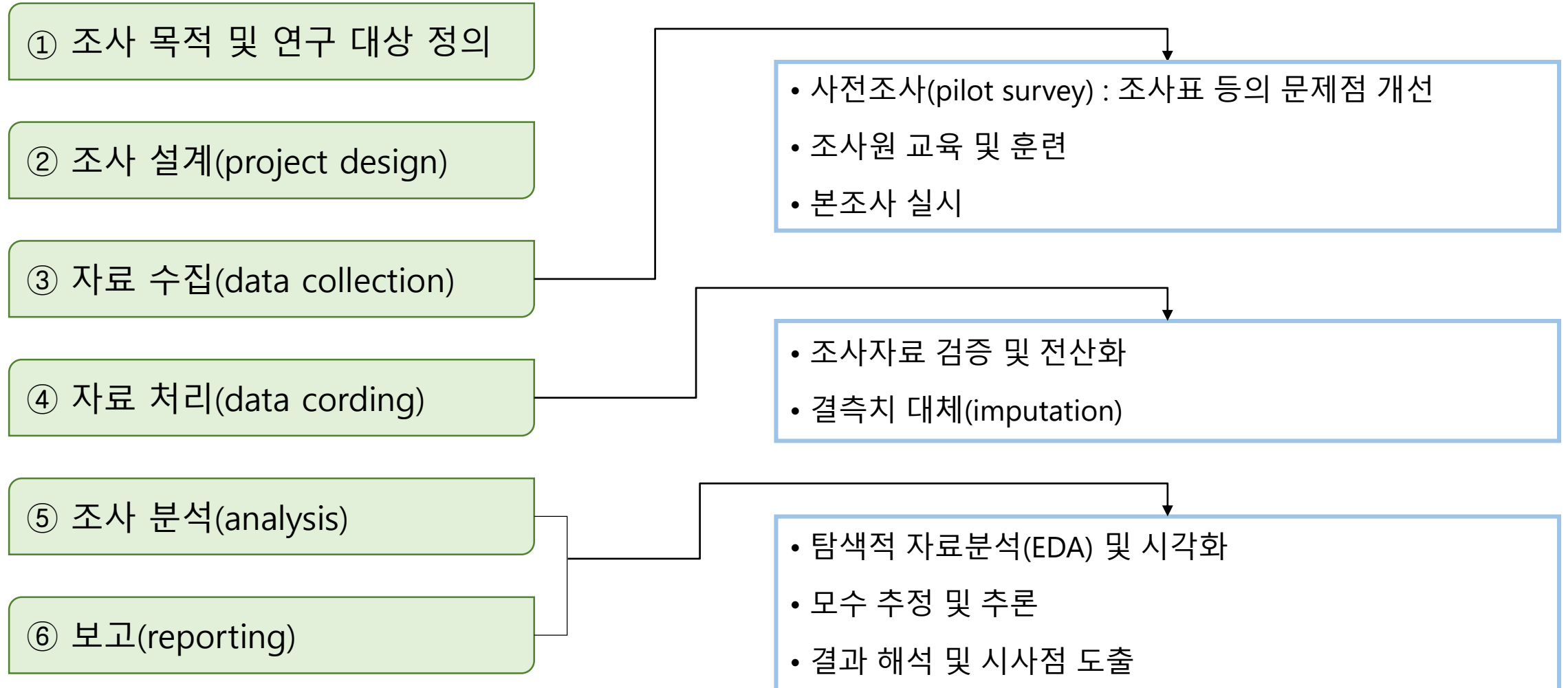
④ 자료 처리(data cording)

⑤ 조사 분석(analysis)

⑥ 보고(reporting)

- 조사형태 결정 : 전수조사 or 표본조사
- 표본추출틀(sampling frame)의 선택  
: 표본추출단위의 명부(list)
- **표본설계(sample design)**  
: 오차의 수준, 표본추출 방법 및 표본크기, 추정방법 등
- 표본추출(sample selection)  
: 본 표본 및 예비표본(조사 불가시 대체할 표본)
- 조사표 설계(questionnaire design)
- 조사방식(면접, 전화, 인터넷 등) 선택

# 조사의 기본 절차(Survey Process)



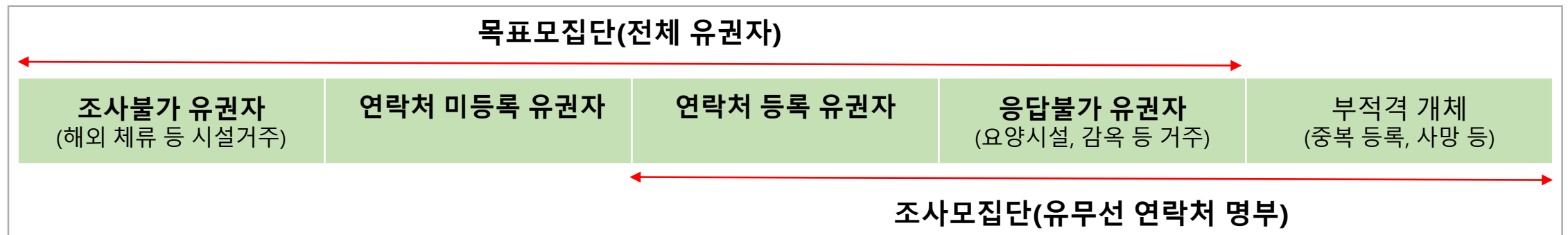
# 목표모집단 vs 조사모집단

## ◆ 목표모집단(target population)

- \* 조사에서 얻어진 추론의 결과를 직접 적용할 수 있는 모집단(ex: 현시점 투표권을 보유한 전체 유권자)
- \* 연구 결과의 일반화 범위를 결정

## ◆ 조사모집단(survey population)

- \* 실제 조사 가능한 범위 내의 전체 집단(ex : 국내 거주중인 연락 가능한 유권자)
- \* 표본추출을 위한 추출틀 및 실제 조사 대상 결정



⇒ 목표모집단과 조사모집단의 차이가 클수록 결과의 대표성과 일반화 가능성이 약화됨

# 전수조사 vs 표본조사

	전수조사 (census)	표본조사 (sample survey)
정의	모집단의 모든 개체를 조사	모집단의 일부 개체만 추출하여 조사
장점	이론상 추정오차 없이 모수를 정확히 계산 가능 (정확성)	비용과 시간 등 자원의 절약 (시의성 확보 및 시계열 추적 용이)
단점	<ul style="list-style-type: none"> <li>조사 누락 또는 중복 문제, 시의성 부족</li> <li>현실적 어려움(조사대상 리스트 확보)</li> </ul>	<ul style="list-style-type: none"> <li>정확성이 표본 및 추정방법에 좌우</li> <li>표본설계 및 추정에 대한 전문성이 요구</li> </ul>
오차	비표본오차	표본오차, 비표본오차
활용	<ul style="list-style-type: none"> <li>결과에 대한 정확성이 더 요구되는 조사</li> <li>모집단이 작고, 접근이 용이한 경우</li> <li>모수의 수준(level)이 중요, 표본조사의 추출틀 제공</li> </ul>	<ul style="list-style-type: none"> <li>결과에 대한 시의성이 더 요구되는 조사</li> <li>전수조사가 불가능하거나 비효율적인 경우</li> <li>모수의 변동량 파악이 중요</li> </ul>
예시	인구주택총조사, 경제총조사	여론조사, 경제활동인구조사

# 사례 1 : 경제총조사

---

## ◆ 통계에 대한 설명자료 : [통계청](#)

### \* 조사 목적

국가 전체 산업에 대한 구조와 분포, 경영실태 등에 관한 사항을 종합적으로 파악하여 경제 및 산업정책 수립 기초자료, 각종 통계의 모집단 자료, 소지역 단위(지역) 통계자료 등으로 활용

### \* 조사 주기 : 5년

### \* 조사 대상

한국표준산업분류 21개 대분류(A~U) 중 T(자가소비 생산활동), U(국제 및 외국기관)를 제외한 산업의 모든 사업체(2020년 조사부터 별도의 사업장이 없는 사업체도 포함)

### \* 목표모집단 및 조사모집단

기업통계등록부(SBR)를 기반으로 조사 기준일 현재 우리나라에서 산업 활동을 수행하고 있는 사업체

### \* 조사 형태 : 전수 + 표본조사 병행(2015년 조사까지는 전수조사)

### \* 자료수집 방법 : (주) 면접조사 / (부) 우편(팩스)조사, 인터넷조사, 전화조사

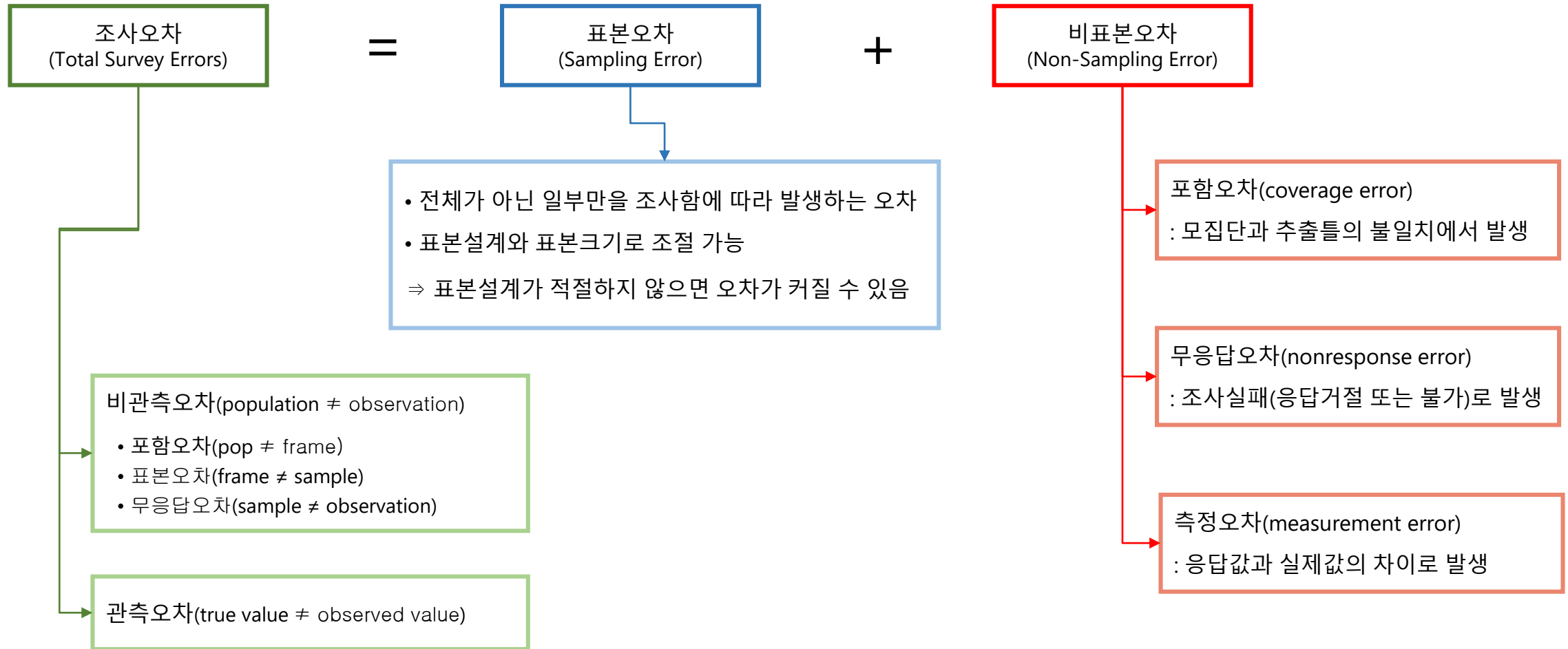
## 사례 2 : 여론조사

---

◆ 통계에 대한 설명자료 : [중앙선거관리위원회](#)

- \* 조사 명 : 한국갤럽 데일리 여론조사
- \* 조사 주기 : 매주
- \* 조사 대상 : 전국 만 18세 이상 남녀
- \* 조사 형태 : 표본조사(표본크기 1,000명)
- \* 표본추출틀 : 무선통신 가입자 리스트로부터 무작위 추출
- \* 자료수집 방법 : 무선전화면접
- \* 응답률 : 14.9%
- \* 표본오차 : 95% 신뢰수준에  $\pm 3.1\%P$

# 조사오차(survey errors)



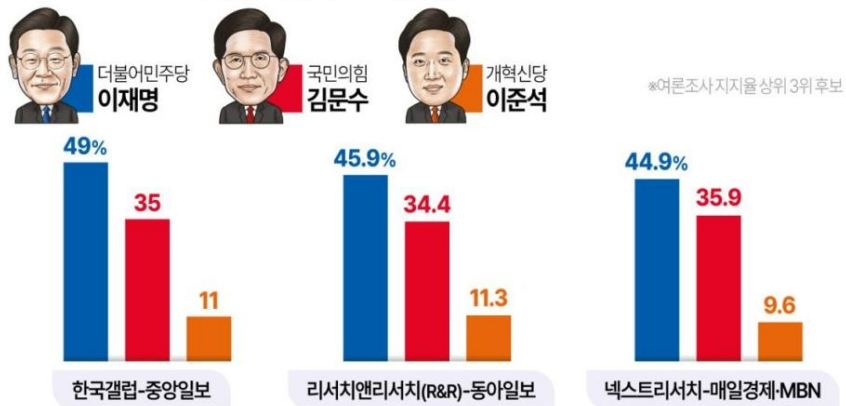


# 조사오차 : 선거 여론조사는 왜 틀릴까?

## ◆ 여론조사기관에서 발표하는 오차범위의 전제

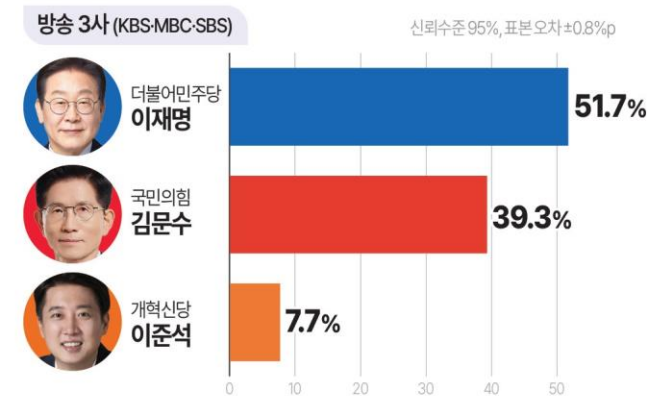
- \* 표본은 무작위로 추출되었음
- \* 표본에 추출된 사람은 모두 응답하였음
- \* 응답자는 모두 자신의 의견을 솔직하게 응답하였음

2025대선 '공표금지 D-1' 주요 여론조사 결과  
5월 28일부터 대통령 선거 여론조사 공표금지



(출처) : 연합뉴스

2025대선 출구조사 결과

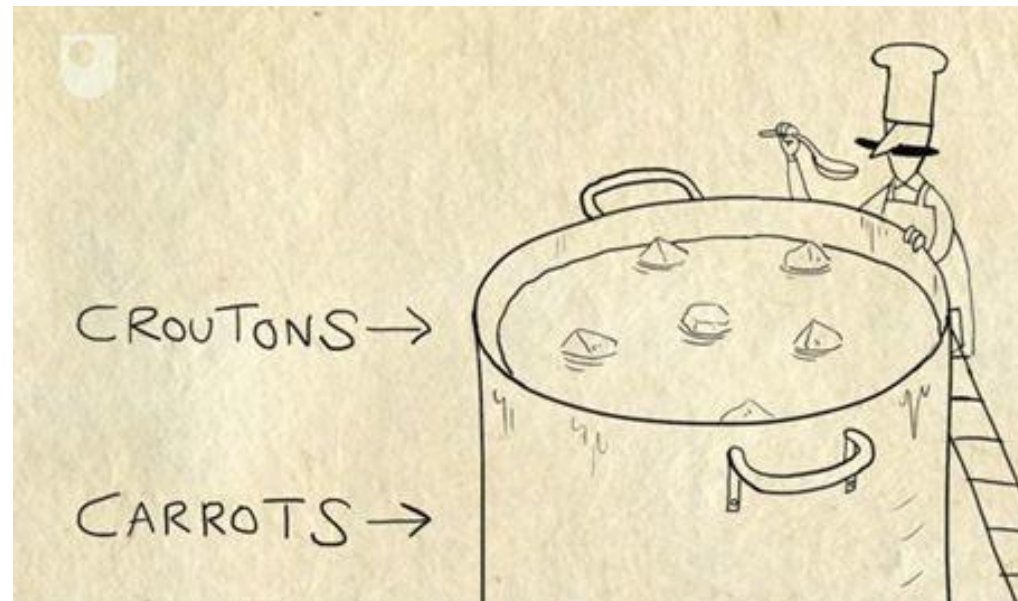


연말뉴스 자료: KBS-MBC-SBS 김영은 기자 = 0eun@yna.co.kr/20250603

# 그럼에도 불구하고...

---

"You **only** need a **single spoon** to sample the soup, provided it is **well stirred**."



Source : <https://www.linkedin.com>