

## 제 1장. 자료의 생성

# 통계학이란?

---

- ▶ 예 : 2017년도 19대 대선 여론조사 결과

후보명	M	A	H	S	Y
지지율	42.4%	18.6%	18.6%	7.3%	4.9%

(2017년 5월 1주차, 리얼미터)

Q1. 전체 유권자 중 누구를 대상으로 조사를 할 것인가?

Q2. 조사한 내용을 어떻게 활용할 것인가?

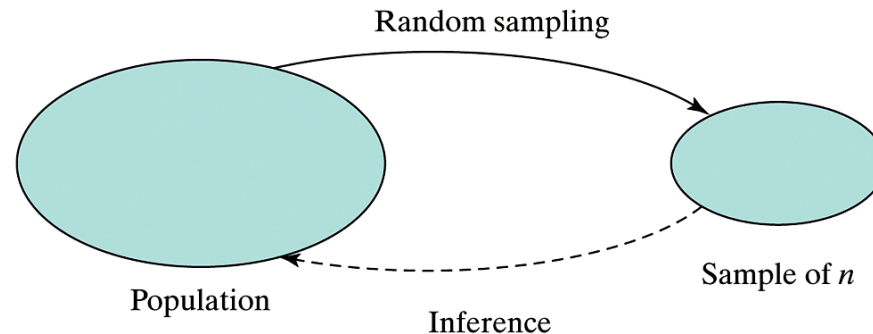
Q3. 조사 결과의 신뢰성은 어떻게 측정할 것인가?

- ▶ **통계학 (statistics)**

: 자료를 수집하고 정리하며, 이를 해석하고 신뢰성 있는 결론을 이끌어내는 방법을 연구하는 학문

# 통계학이란?

- ▶ 관심과 추측의 대상이 되는 전체 중에서 일부분에 대한 자료를 수집하고 이를 통하여 전체에 대한 합리적인 결론을 이끌어 내는 과정



- ▶ 추출단위 (sampling unit) : 전체를 구성하는 각 개체
- ▶ 특성값 (characteristic) : 각 추출단위의 특성을 나타내는 값
- ▶ 모집단 (population) : 모든 추출단위의 특성값을 모아놓은 것  
: 관심과 추측의 대상이 되는 전체
  - 유한 모집단 (finite population)
  - 무한 모집단 (infinite population)
- ▶ 표본 (sample) : 실제로 관측한 것들의 모임  
: 모집단의 부분집합

# 자료의 수집

---

- ▶ **조사(survey)** : 관찰 대상을 통해 관심 있는 특성을 관찰하는 것
  - **전수조사(census)**
    - 모집단의 모든 단위들을 다 조사하는 방법
    - 정확하지만 시간과 비용이 많이 듦
    - 예 : 인구주택총조사
  - **표본조사(sample survey)**
    - 모집단의 부분집합인 표본만을 조사하는 방법
    - 시간과 비용을 절약할 수 있으나 표본에 따라 정확성이 떨어질 수 있음
    - 표본이 모집단을 잘 대표할 수 있도록 해야 함
    - 예 : 경제활동인구조사 등 대부분의 조사
  
- ▶ **실험(experiment)** : 원하는 목적의 데이터를 얻는데 가장 효과적인 방법
  - 관심의 대상 이외의 요인을 통제하는 것이 반드시 필요함
  - 실험 전 미리 실험 설계를 하는 것이 매우 중요함

# 단순랜덤추출

---

## ▶ 단순랜덤추출법 (simple random sampling)

- : 모집단의 속성을 잘 대표할 수 있도록 표본을 '골고루' 뽑는 방법
  - : 유한모집단에서  $n$ 개의 추출단위로 구성된 모든 부분집합들이 표본으로 선택될 확률이 같도록 설계된 표본 추출방법
  - : 개념적으로 간단하며 통계이론 전개의 기본 틀로 사용
  - : 대상 집단의 모든 구성요소에 대해 동등한 확률을 부여하여 표본 추출  
("simple = equal probability")
  - : 조사자의 주관이 개입되지 않도록 난수표나 추첨을 통해 표본을 랜덤으로 추출
- 단순랜덤 복원추출 (SRS with replacement)
  - 단순랜덤 비복원추출 (SRS without replacement)

# 자료의 종류

---

- ▶ 자료의 특성에 따라 적용할 수 있는 통계분석 방법들이 다르고, 사용할 수 있는 그래프 또는 요약 방법도 달라지기 때문에 자료의 특성을 분류하는 것이 중요함
- ▶ **연속형 자료 (continuous data)**
  - : 구간 내의 연속적인 값을 갖는 자료들 (키, 몸무게, 전구의 수명)
- ▶ **이산형 자료 (discrete data)**
  - : 이산적인 값을 갖는 자료들
  - (1) 계수형 자료 (counting data)
    - : 숫자로 표현 가능한 이산형 자료 (불량품의 개수, 전화의 횟수, 자동차사고 건수)
  - (2) 범주형 자료 (categorical data)
    - : 숫자로 표현할 수 없는 이산형 자료 (성별, 직업, 교과목)
    - : 숫자로 표현하더라도 기호의 개념이 됨

## 표본조사시 유의사항

---

- ▶ 표본은 어느 한 쪽으로 치우치지 않고 모집단을 골고루 잘 대표할 수 있어야 함
- ▶ 모집단의 정확하고 완전한 리스트가 필요
  - : 표본의 치우침 현상 방지
- ▶ 무응답(non-response)의 방지
  - : 응답자 그룹과 무응답 그룹간의 비교 분석이 필요
- ▶ 철저한 조사자의 훈련 및 감독
- ▶ 정확한 설문지 작성