

# 제1장 통계학의 내용

## ▼ 01 통계학의 소개

- **통계학(Statistics)** : 사회 현상 or 자연 현상에 내재한 불확실성을 수치로 요약
- **(1) 기술통계학과 추측통계학**

기술 통계학 (descriptive statistics)	- 자료를 분석, 제시 - 두 시점의 자료를 비교하는 것은 자료를 단순히 수집/정리/요약/분석 - 방대한 자료를 일목요연하게 표현 - 자료의 특성을 수치로 요약 - 추측 통계학의 사전 단계 - 불확실한 미래에 대한 예측과 의사 결정을 합리적으로 수행	
추측 통계학 (inferential statistics)	- ex) 품질검사/미래의 예측/대륙붕시추에 대한 의사결정/단당류와 항암작용의 관계 - 완전한 정보가 아니고 단지 부분적인 정보에 근거하여 판단을 내리는 것 - 표본의 결과로부터 모집단의 특성을 추측하고 일반화or추론 하는 과정	

- **(2) 통계와 통계학**
  - 통계 : 관심의 대상이 되는 집단의 관찰로 부터 얻은 수치적 자료(numerical data) or 요약
  - 통계학
    - 통계 자료를 분석 > 불확실한 상황에서 판단을 객관화 or 합리화 도움
    - 관심의 대상이 되는 자료를 수집/정리/요약/분석 > 의사결정에 도움

- **(3) 모집단과 표본**

- 모집단(poulation) : 관심의 대상이 되는 모든 잠재적 관찰 대상의 집합

유한 모집단 (finite population)	- 모집단을 구성하는 개체들이 한정
무한 모집단	- ex 제품의 수명 > 0보다 큰 실수값

- 표본(sample)
  - 통계적 처리를 위해 모집단에서 추출된 모집단의 일부
  - 모집단의 특성을 잘 반영해야함.

- **(4) 모수와 통계량**

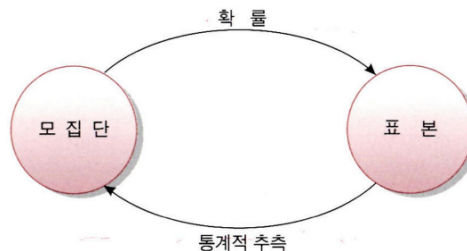
- 모수(population parameter) : 모집단의 특성을 나타냄 ex)평균/분산/비율

표 1-1 모집단과 모수의 구별

모 집 단	모 수
선거권을 갖고 있는 대학생	선거명부에 등록된 대학생의 비율
대한민국 국민	남녀의 성비율
지하철 이용객	지하철 이용객의 평균 만족도
모 은행의 통장보유자	10월까지 실명확인을 마친 고객의 수
한달동안 복권을 구입한 사람	한달동안의 1인당 평균 복권 구입량

- 통계량(sample statistic) : 표본의 특성
- (5) 확률과 통계적 추측
  - 확률(probability) : 모집단의 특성을 알고 표본의 결과에 대한 가능성을 계산할 때 적용이 되는 수학적 언어
  - 통계적 추측(statistical inference) : 표본의 결과를 근거로 모집단의 특성을 파악하는 행위

그림 1-1 모집단과 표본의 관계



- (6) 표본오차와 비표본오차
  - 전수 조사(census) : 모집단 전체에 관한 완전한 조사
  - 표본조사(sampling survey) : 표본을 추출하여 관측하는 과정
  - 표본오차(sampling error)
    - 표본의 결과를 근거로 모집단 전체의 특성을 추론하는 과정에서 발생하는 오차
    - 전수조사의 경우 표본오차X
    - 표본의 크기가 증가함에 따라 감소
  - 비표본오차(nonsampling error)

- 표본조사 전과정에서 발생할 수 있는 오차
- 원인 : 조사 목적의 불분명함, 모집단의 비합리적 설정, 부정확한 측정, 관찰의 오류, 누락, 오기, 부정직 등
- 표본의 크기가 커짐에 따라 증가O
- 전수조사에서도 발생

## ▼ 02 표본 추출의 방법

### ▼ 표본 추출 방법 요약

- 표본추출
  - 판단표본 추출
  - 확률표본추출
    - 단순무작위표본추출
    - 층화표본추출
    - 군집표본추출

### ▼ 2.1 판단표본추출 10



판단표본추출(judgement sampling) : 전문가의 의견에 따라서 표본을 추출

- 객관적인 분석 곤란
- 표본의 결과를 이용하여 모집단의 특성을 추정하는 것이 불확실한 결점

### ▼ 2.2 확률표본추출 10

- 확률표본추출(probability sampling)
  - 표본을 추출하기 전에 모집단을 구성하는 기본단위가 표본으로 추출될 확률을 일정하게 할당하는 방법
  - 객관적인 분석 가능
- (1) 단순무작위표본추출(simple random sampling)
  - 일정 크기의 모든 표본조합(sample combination)이 표본으로 추출될 확률을 같게함

- 모집단의 기본단위가 표본에 포함될 확률을 같게 하여 표본을 추출하는 방법



ex) 10명의 선수를 가진 농구팀.

- 각 선수가 경기 당 올리는 평균 득점 > 3명의 평균득점으로 추정
- 표본의 크기 3.
- 표본의 수 :

$${}_{10}C_3 = \frac{10!}{3!7!} = 120\text{개}$$

- 표본조합이 표본으로 추출될 확률 = 1/120  
각 선수가 표본에 포함될 확률 = 1/10

- 난수(randomnumber) = 균등분포(uniform distribution)로 발생하는 숫자

## • (2) 층화표본추출

- 층(strata) : 몇 개의 동질적인 집단으로 모집단을 구분해 놓은 것을 의미
- 층화표본추출(stratified sampling)
  - 모집단이 몇 개의 층으로 나누어 질 때 각 층으로부터 표본을 추출하여 전체 표본을 구성하는 방법
  - 기본단위의 특성이 매우 다양하고 사전에 모집단의 구조를 잘 알고 있을 때 사용
  - 각 층으로부터 추출된 표본의 결과에 가중치를 적용, 모집단 특성에 대한 추정치를 계산

## • (3) 군집표본추출(Cluster sampling)

- single-stage sampling : 1단계 추출 - 추출된 군의 기본단위를 전부 표본에 포함
  - ex ) 대학교 몇 개 추출
- two-stage sampling : 2단계 추출 - 다시 한번 표본을 추출하는 경우
  - ex ) 학과 몇 개 추출

- multi-stage sampling : 다단계 추출 - 두 단계 이상의 추출이 있는 경우
  - ex) 학생을 표본으로 추출
- 효과적일 때
  - 군내에는 기본단위의 특성 차이가 큼
  - 군간에는 차이가 작을 때
  - 비용 및 시간을 절약
- (4) 체계적 표본추출(systematic sampling)
  - 모집단의 크기(N)표본(n).
  - 순서
    - 모집단의 기본단위를 순서대로 1~N까지 번호 할당
    - 표본추출간격(sampling interval)  $k = N/n$
    - 첫째 구간에서 무작위로 기본단위 하나를 추출
    - if 번호=a > 표본은 번호가 a,a+k,a+2k,a+3k...인 기본단위들을 추출하여 구성

✓ 100,000명 중 200명의 표본을 추출

$$\text{표본추출구간은 } k = \frac{100,000}{200} = 500$$

- 500명마다 1명씩 추출하게 되는 것.

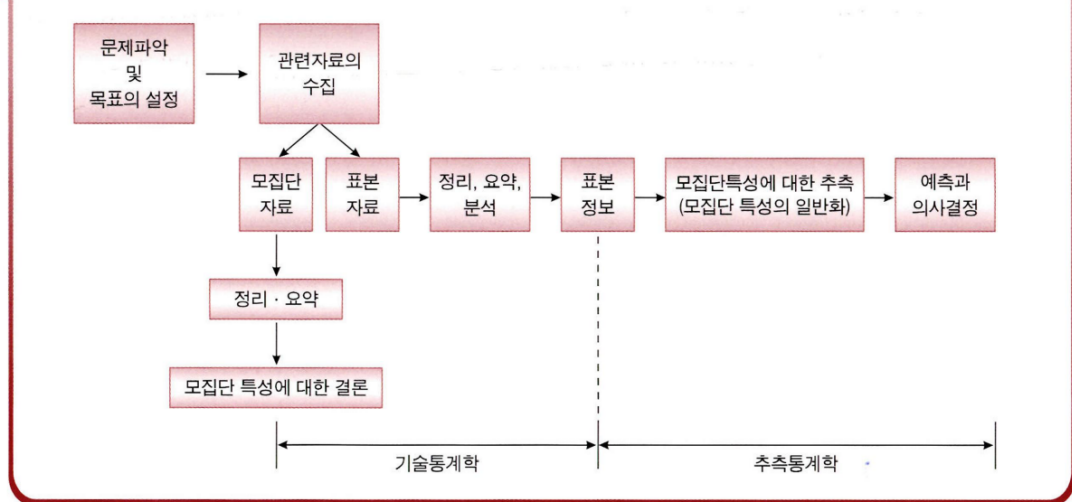
- 모집단의 순서에 따른 주기성이 표본추출간격과 일치하는 경우에는 동일한 유형의 기본단위들로만 표본이 구성되어 모집단의 실제 특성을 반영하지 못하는 문제가 발생할 수 있음.

### ▼ 03 통계학과 의사결정

#### 통계적 의사결정(statistical decision making)

- 표본정보를 기초로 모집단의 특성을 추측 > 예측or의사결정 수반>부분적인 정보를 근거로 예측or의사결정을 내리는 것

그림 1-3 통계학의 문제해결절차



• 통계적 의사 결정의 절차 요약

- 1) 당면 문제를 정확하게 파악하여 목표 설정
- 2) 필요한 자료 수집/정리/요약/분석 > 표본의 특성 파악
- 3) 표본정보를 기초로 모집단 특성에 대한 추측
- 4) 미래에 대한 예측과 의사 결정을 수행