



1

## 제1장 통계학의 개념

- 1.1 의사결정과 통계적 사고
- 1.2 통계학이란?
- 1.3 모집단과 표본
- 1.4 샘플링의 개념
- 1.5 데이터의 종류
- 1.6 기술통계와 추측통계

2/17

2

## 1.1 의사결정과 통계적 사고



- 문제해결을 위한 **의사결정 프로세스**

- ① 핵심 이슈 정리
- ② 조사 대상 선별
- ③ 필요 데이터 규명, 실험/조사 방법 수립
- ④ 데이터 수집, 정리, 통계 처리 → 정보
- ⑤ 중요한 요인(factor) 발견
- ⑥ 통계적 모형 개발, 추정
- ⑦ 적합성 검증, 모형 최적화 → 해답
- ⑧ 결론 도출, 의사결정 제안

3/17

3

## 1.2 통계학이란?



- [정의 1-1] **통계학(statistics)**

- 불확실한 상황 하에서 데이터에 근거하여 과학적인 의사결정을 하기 위한 이론과 방법의 체계

- 어원

- 라틴어 **statisticus**(확률), **statisticum**(상태)
- 이탈리아어 **statista**(나라, 정치가)

4/17

4

## 1.2.2 통계학의 활용분야



- 데이터의 수집, 분류, 표현, 분석, 추론, 해석
- 공학, 농업, 생명과학, 환경과학, 경제학, 경영학, 산업연구, 품질보증, 시장조사, 여론조사, ...
- 통계학자의 역할
  - 필요한 데이터의 형태
  - 데이터를 수집하는 방법
  - 문제에 대한 최선의 답을 구하기 위한 분석방법

5/17

5

## 1.2.3 통계학의 연구분야



- 통계적 이론과 방법론 개발
  - 확률론, 확률과정론, 수리통계, 점근적 추론,
  - 선형/비선형모형, 회귀분석, 실험계획, 통계계산,
  - 베이지안 추론, 비모수추론, 시계열 분석 등
- 통계적 방법론 응용
  - 생물통계, 생명정보학, 품질관리, 신뢰성분석,
  - 공학통계, 계량경제학, 금융통계, 보험통계, 기상통계

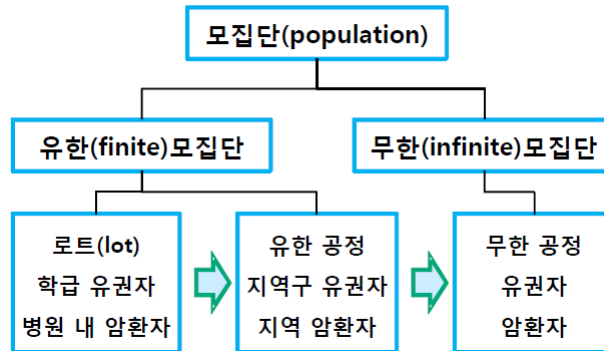
6/17

6

## 1.3 모집단과 표본

### [정의 1-2] 모집단(population)

- 통계분석의 연구대상이 되는 **모든 개체들의 집합**, 올바른 의사결정을 하기 위해 관심을 갖고 **연구해야 할 대상**



7/17

7

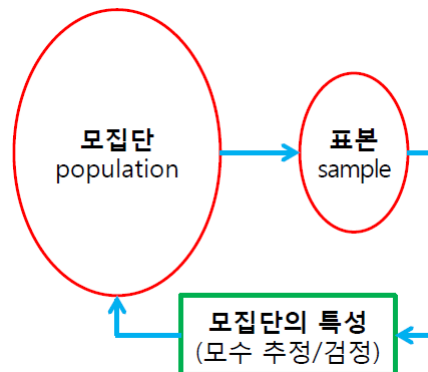
## 1.3 모집단과 표본

### [정의 1-3] 모수(parameter)

- 모집단의 특성을 나타내는 수치로서, 올바른 의사결정을 위해 특별히 관심을 갖는 모수만이 연구의 대상이 됨.
- ✓ 평균, 분산, 표준편차, 비율 등

### [정의 1-4] 표본(sample)

- 모집단의 특성을 파악하기 위해 모집단으로부터 **일정한 규칙에 의해 추출한 모집단의 부분집합**.



8/17

8

## 1.3 모집단과 표본

- 통계적 추론(inference)

- 모집단의 특성(평균, 분산, 비율 등)을 추측하는 것.  
추정(estimation), 가설검정(hypothesis test)

- 통계량(statistic)

- 모집단의 특성(평균, 분산, 비율 등)을 추측하기 위해 사용하는 표본의 함수

- 추정량(estimator)

- 모수를 추정하는데 사용하는 통계량.  
표본평균, 표본분산, 표본비율 등

$$\bar{X} \Rightarrow \mu$$

$$S^2 \Rightarrow \sigma^2$$

$$\frac{X}{n} \Rightarrow p$$

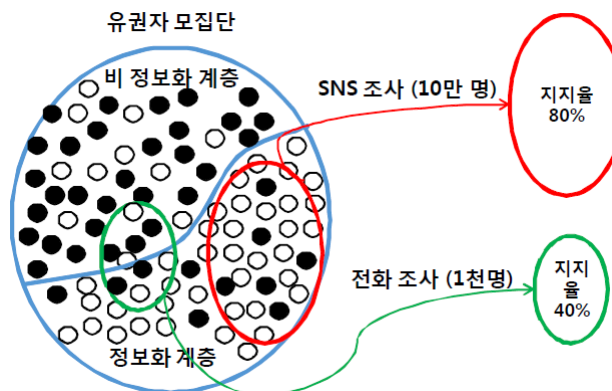
9/17

9

## 1.4 샘플링의 개념

[정의 1-5] 샘플링(sampling)

- 모집단에 대한 올바른 의사결정을 내리기 위해 모집단으로부터 표본을 추출하는 일련의 과정



10/17

10

## 1.4.1 샘플링의 필요성

- 표본조사(sample survey)

- 모집단에서 표본을 추출하는 샘플링에 의한 조사 방식

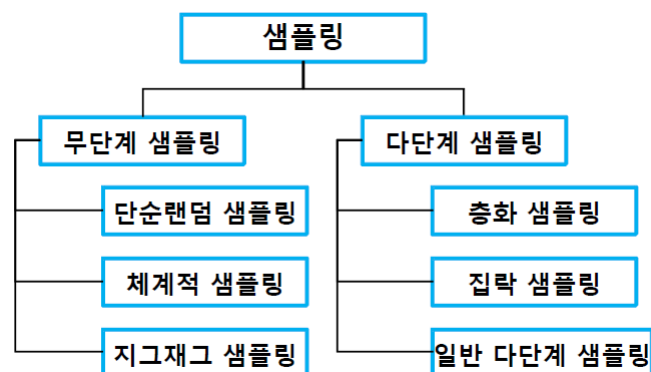
- 샘플링의 필요성

- (1) 현실적 제약
  - ✓ 무한 모집단, 무의미, 파괴 검사, ...
- (2) 비용 절감
  - ✓ 인건비, 분석비용, 부대비용, ...
- (3) 시간 절약
  - ✓ 신속한 조치, 시간적 제약, ...
- (4) 품질 측면
  - ✓ 피로도 고려, 검사 실수, ...

11/17

11

## 1.4.2 샘플링 방식



[정의 1-6] **확률표본(random sample)**

- 모집단으로부터 랜덤하게 샘플링하여 독립적이고 동일한 분포 (identically independent distributed)를 따르도록 형성된 표본

12/17

12

## 1.4.2 샘플링 방식

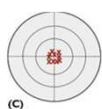
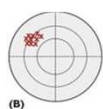
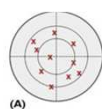
- 50개 들이 20개 상자로부터 100개의 표본을 샘플링
  - ① **단순랜덤 샘플링** 1~1000 사이에서 랜덤하게 100개 선택
  - ② **체계적 샘플링** 1~10 사이에서 랜덤하게 1개 선택 → 7  
→ 7, 17, 27, ..., 977, 987, 997 등 100개 선택
  - ③ **지그재그 샘플링** → 7, 14, 27, ..., 974, 987, 994 등 100개 선택
  - ④ **2단계 샘플링** 상자 5개를 랜덤 샘플링, 각 상자에서 20개씩 랜덤 샘플링
  - ⑤ **층화 샘플링** 모든 20개 상자에 대하여 각 상자에서 5개씩 랜덤 샘플링
  - ⑥ **집락 샘플링** 2개 상자를 랜덤 샘플링, 50개 모두 선택

13/17

13

## 1.4.3 샘플링과 오차

- **오차(error)** : 모수의 참값과 측정값과의 차이
- **오차의 유형**  $x = \mu + s + r + m$ 
  - 측정오차 : 측정계기의 부정확, 측정기술 부족
  - 분석오차 : 분석장비의 부정확, 분석기술 부족
  - 샘플링오차 : 치우친 샘플링
  - 축분오차 : 큰 표본에서 측정용 표본을 샘플링할 때 치우침
- **오차의 검토**
  - ① 신뢰성(reliability) : 정밀도, 정확성 동시 유지할 때 가능
  - ② 정밀도(precision) : 표준편차, CV가 적을수록 높음
  - ③ 정확성(accuracy) : 회수율, 상대오차, 치우침, t 검정



- (A) 정확성 : 높음, 정밀성 : 낮음
- (B) 정확성 : 낮음, 정밀성 : 높음
- (C) 정확성 : 높음, 정밀성 : 높음

14/17

14

## 1.5 데이터의 종류

- 데이터 척도(Scale)의 유형

척도 유형	특성	사례
a. 명목척도 (nominal)	순서, 크기 등의 의미 없음	교과목번호, 차번호, 성별, 주민번호 뒷자리, 복권번호
b. 순서척도 (ordinal)	순서의 의미는 있으나, 간격의 의미는 없음	직급, 계급, 우선순위, 선호도
c. 구간척도 (interval)	순서와 간격의 의미는 있으나, 비율(혹은 절대 0)의 의미는 없음	대기표번호, 화씨온도, 지능지수, 표준점수
d. 비율척도 (ratio)	순서, 간격, 비율(혹은 절대 0)등의 의미가 모두 있음	농도, 무게, 길이, 압력, 속도, 지지율

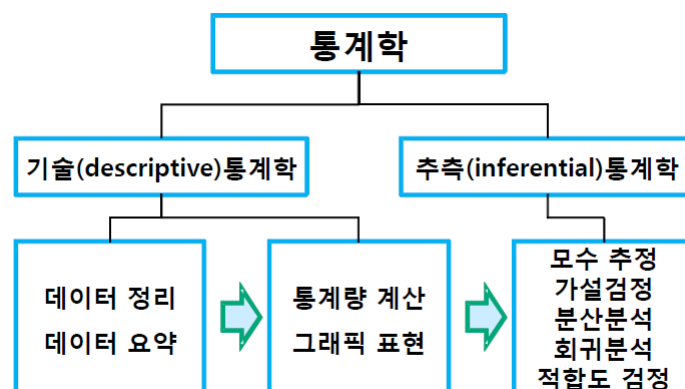
- (1) 절대온도 [ d ]      (2) 화씨온도 [ c ]      (3) 기압 [ d ]      (4) 전화번호 [ a ]  
 (5) 사변 [ a+b+a ]      (6) 학점(GPA) [ d ]      (7) 수능 등급 [ c ]

15/17

15

## 1.6 기술통계와 추측통계

- 통계학의 구성 체계

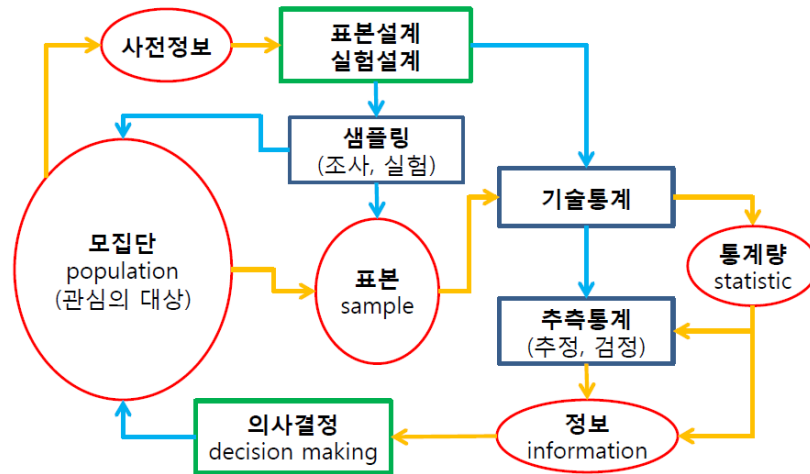


16/17

16



- 통계적 의사결정 절차



17/17