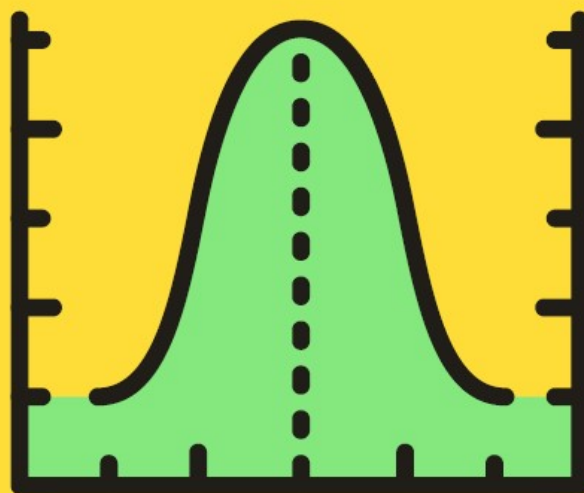


2020년 2학기

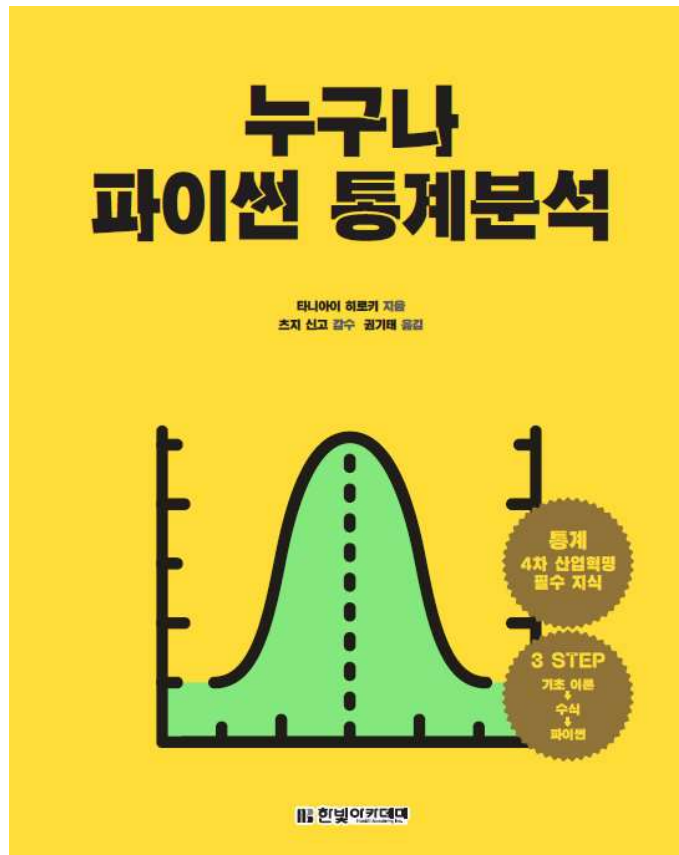
데이터 통계

이동훈
한국산업기술대학교 컴퓨터공학부



누구나 파이썬 통계분석

교재 정보



- 도서명: 누구나 파이썬 통계분석
- ISBN: 979-11-5664-488-0 93310
- 저자: 타니아이 히로키
- 감수자: 츠지 신고
- 역자: 권기태
- 출판사: 한빛아카데미(주)
- 페이지: 384p
- 정가: 26,000원

소스코드 다운로드

https://www.hanbit.co.kr/store/books/look.php?p_code=B2049588591

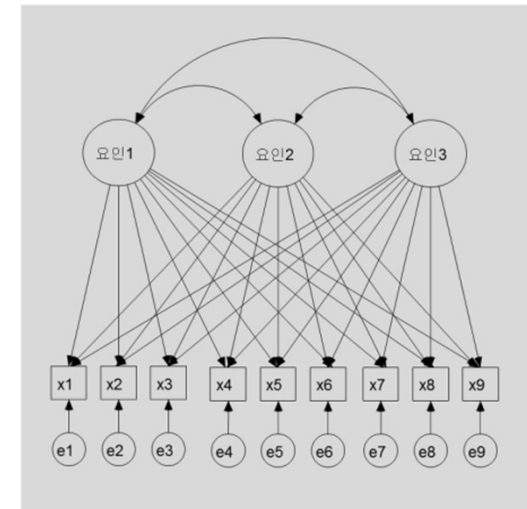
강의 개요

- 과목명 : 데이터통계
- 강의 내용
 - 강의자료는 기본교재를 중심으로 진행.
 - 프로그램 실습은 파이썬코드로 진행.
 - 기본교재 이외의 확률과 통계 관련 자료 및 동영상자료를 제공 → E-Class.
 -
- 성적 관리
 - 중간(30%) + 기말(30%) + 과제(30%) + 출석(10%)
 -

통계학의 개념 - 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



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)

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

통계학의 연구분야

① 통계적 이론과 방법론 개발

- 확률론, 확률과정론, 수리통계, 점근적 추론,
- 선형/비선형모형, 회귀분석, 실험계획, 통계계산,
- 베이지안 추론, 비모수추론, 시계열 분석 등

② 통계적 방법론 응용

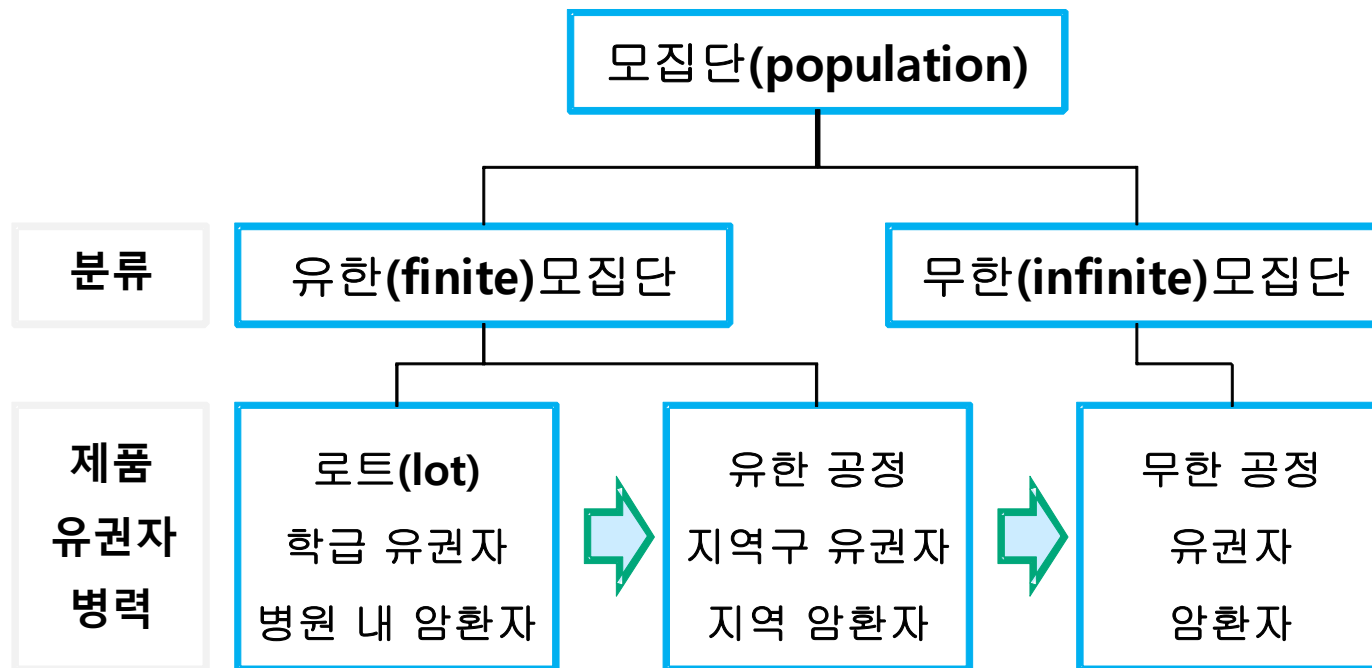
- 생물통계, 생명정보학, 품질관리, 신뢰성분석,
- 공학통계, 계량경제학, 금융통계, 보험통계, 기상통계

| 순위 | 저널명 | 인용지수 | 5-Yr IF | 분야 |
|----|--|-------|---------|------|
| 1 | Statistical Methods in Medical Research | 4.472 | 3.734 | 의학 |
| 2 | ECONOMETRICA | 3.889 | 5.758 | 경제학 |
| 3 | Journal of Statistical Software | 3.801 | 8.309 | 컴퓨터 |
| 4 | Journal of Royal Statistical Society - Series B | 3.515 | 6.585 | 통계학 |
| 5 | Statistical Science | 2.738 | 4.094 | 통계학 |
| 6 | BIOSTATISTICS | 2.649 | 2.853 | 생물학 |
| 7 | Multivariate Behavioral Research | 2.477 | 4.084 | 심리학 |
| 8 | Chemometrics and Intelligent Laboratory Systems | 2.321 | 2.77 | 화학 |
| 9 | Journal of Business & Economic Statistics | 2.241 | 2.928 | 경제학 |
| 10 | The Annals of Statistics | 2.18 | 3.975 | 통계학 |
| 11 | British Jour. of Mathematical & Statistical Psychology | 2.167 | 1.954 | 심리학 |
| 12 | Stochastic Environmental Research & Risk Assessment | 2.086 | 2.348 | 환경과학 |
| 13 | Fuzzy Sets and Systems | 1.986 | 2.496 | 경영과학 |
| 14 | Journal of the American Statistical Association | 1.979 | 3.052 | 통계학 |
| 15 | Probabilistic Engineering Mechanics | 1.855 | 1.722 | 역학 |
| 16 | Statistics in Medicine | 1.825 | 2.609 | 약학 |
| 17 | TECHNOMETRICS | 1.814 | 2.468 | 공학 |
| 18 | Journal of Computational Biology | 1.737 | 1.818 | 생물학 |
| 19 | Journal of Royal Statistical Society – Series A | 1.643 | 2.296 | 통계학 |
| 20 | Statistics and Computing | 1.623 | 1.915 | 컴퓨터 |
| 21 | Spatial Statistics | 1.605 | 1.605 | 환경과학 |
| 22 | BIOMETRICS | 1.568 | 1.882 | 생물학 |
| 23 | Probability Theory and Related Fields | 1.532 | 1.7 | 확률응용 |
| 24 | ENVIRONMETRICS | 1.514 | 1.502 | 환경과학 |

모집단과 표본

[정의 1-2] 모집단(population)

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



1.3.3 모집단과 표본

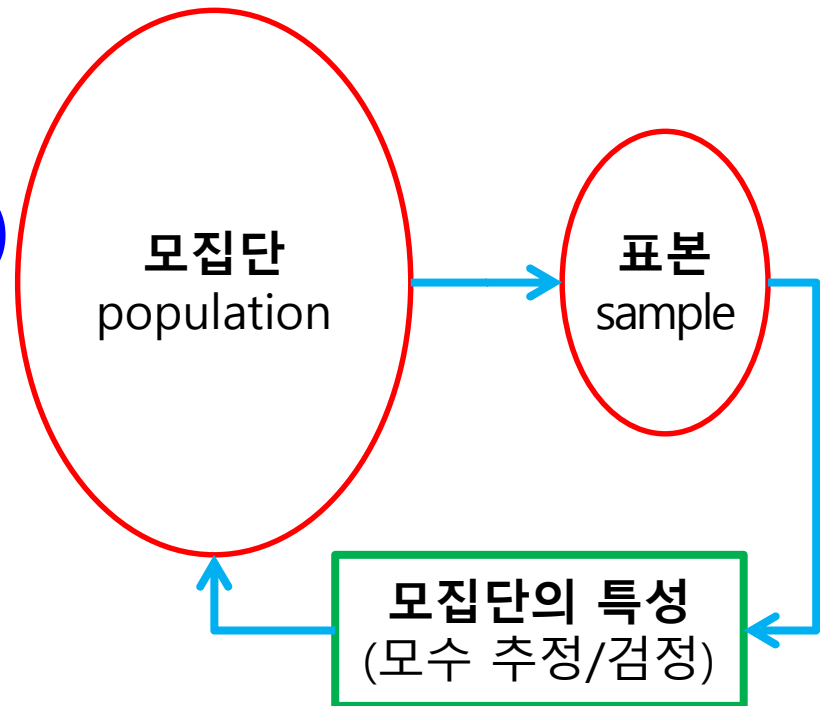
[정의 1-3] 모수(parameter)

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.



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

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

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

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

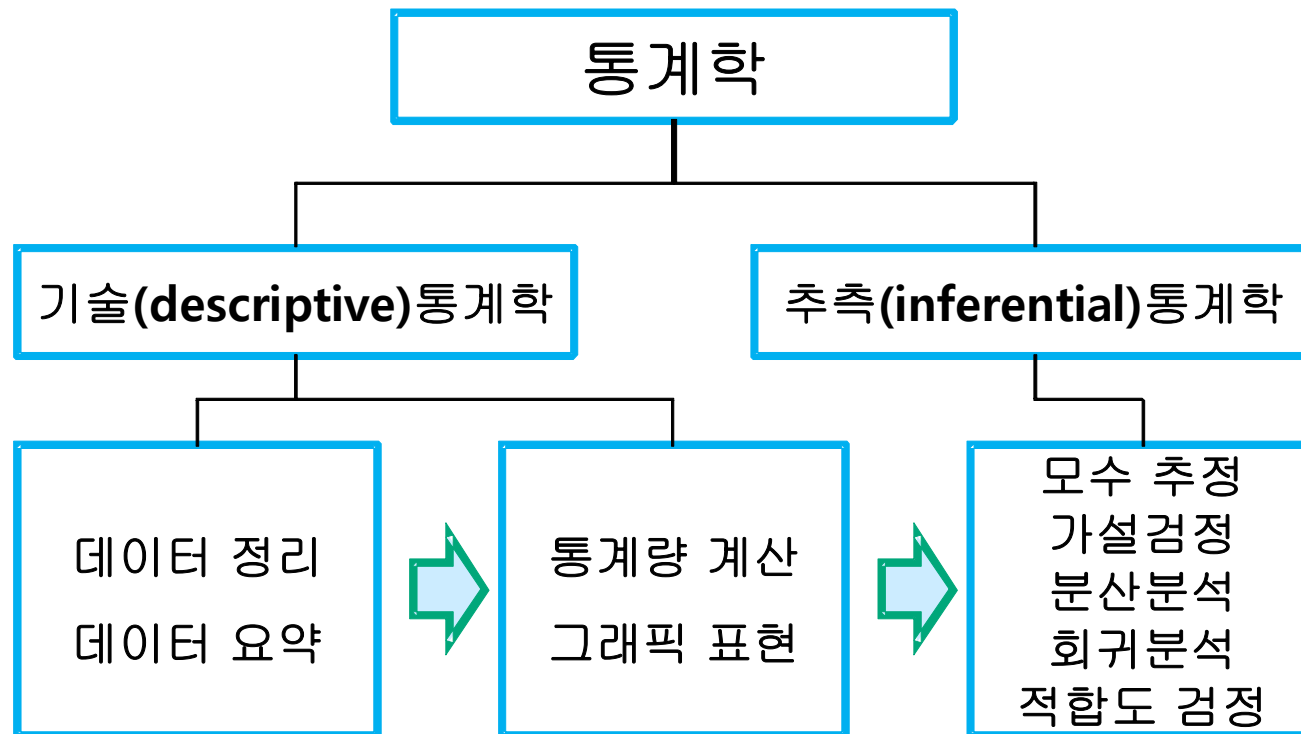
데이터의 종류

➤ 데이터 척도의 유형

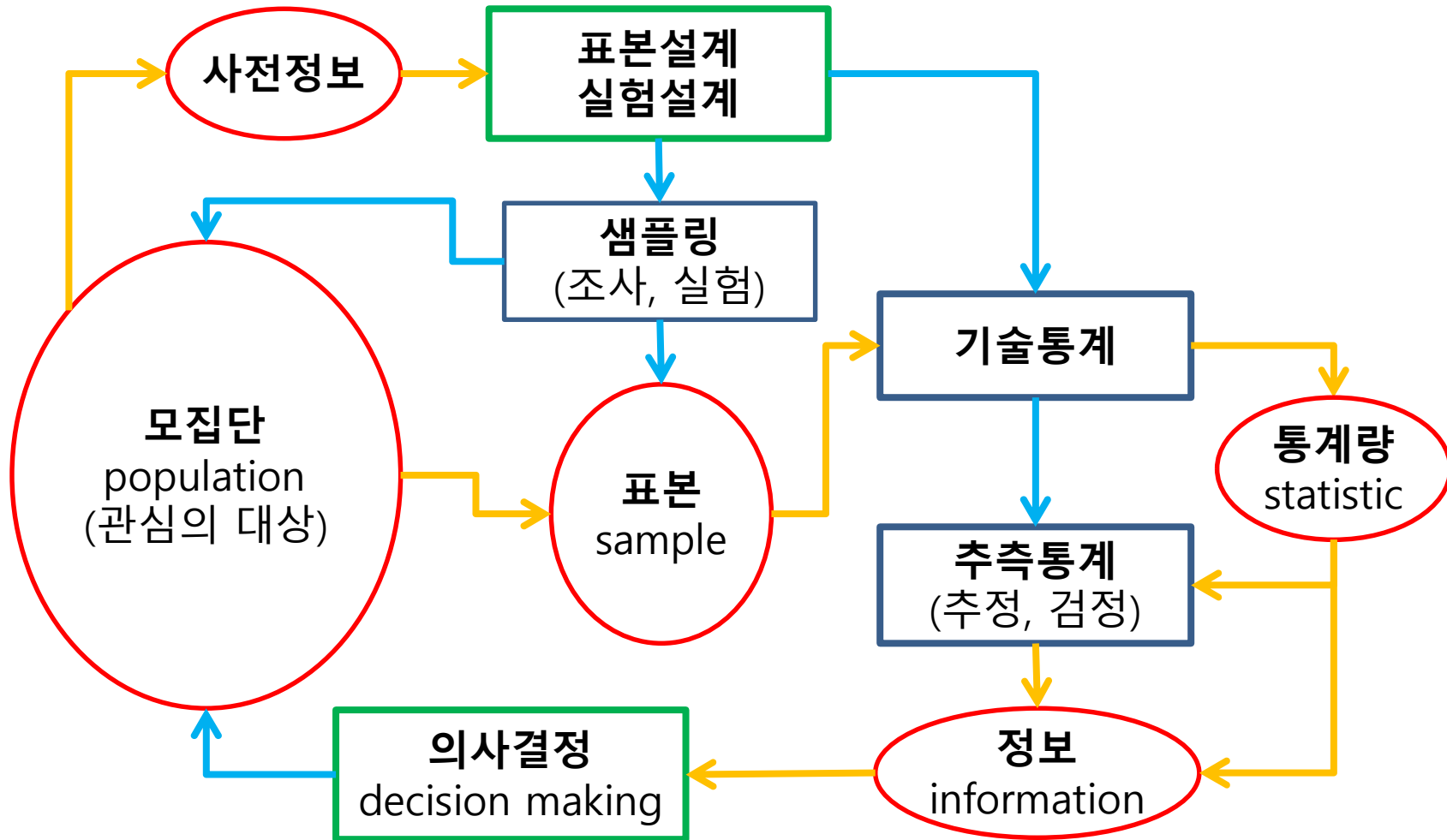
| 척도 유형 | 특성 | 사례 |
|--------------------|--|-------------------------------------|
| 명목척도 (nominal) | 순서, 크기 등의 의미 없음 | 교과코드, 차번호, 성별, 주민번호 뒷자리, 복권번호 |
| 순서척도 (ordinal) | 순서의 의미는 있으나, 간격의 의미는 없음 | 직급, 계급, 우선순위, 선호도 |
| 구간척도 (interval) | 순서와 간격의 의미는 있으나, 비 율(혹은 절대 0)의 의미는 없음 | 대기표번호, 화씨온도, 지능지수, 표준점수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

기술통계와 추측통계

➤ 통계학의 구성 체계



➤ 통계적 의사결정 절차



Thank You!