

추론통계: 모집단과 표본 및 가설

숙명여자대학교 경영학부 오중산

모집단과 표본

- 모집단(population)이란?

- ◆ 데이터를 획득할 사례(case) 전체를 모집단이라고 함

- 예1: 국내 200개 4년제 대학 재학생 150만 명
 - 예2: 18세 이상 유권자 4,000만 명
 - 모집단 크기는 N 으로 표기함

- ◆ 모집단을 대상으로 변수의 측정값, 즉 데이터 획득은 거의 불가능함

- 시간과 비용 측면에서 효율성이 떨어지고, 누락되는 사례가 존재함

모집단과 표본

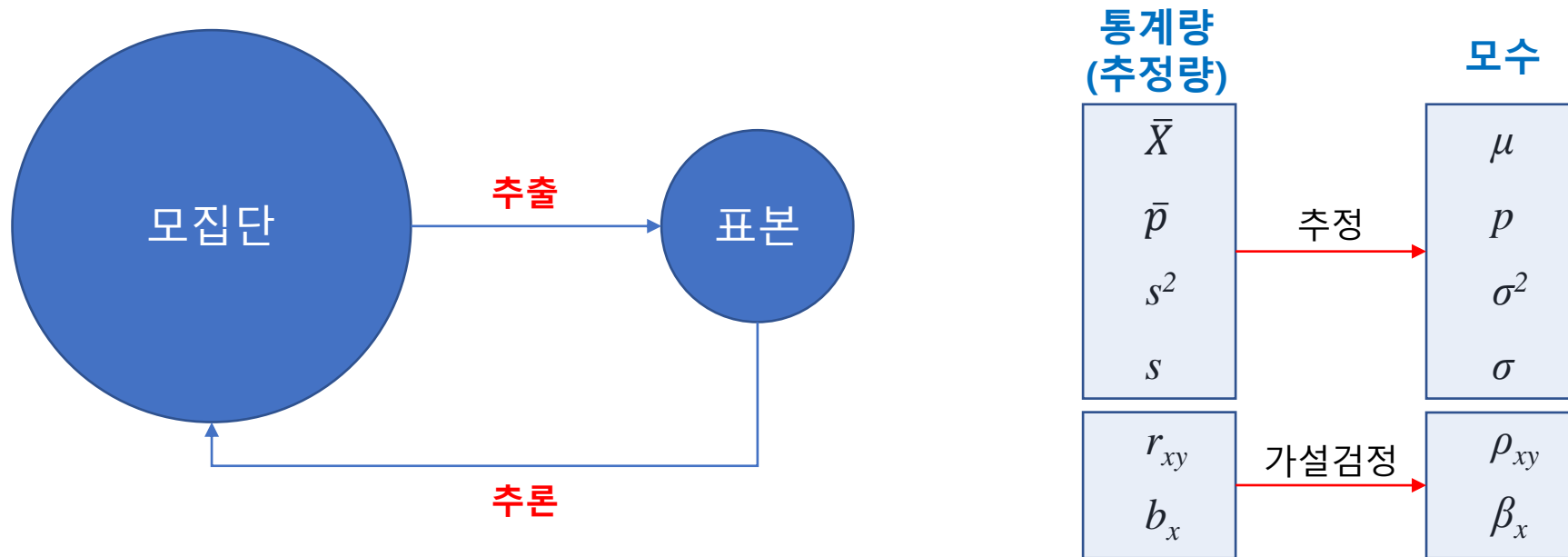
- 표본(sample)이란?

- ◆ 모집단의 일부이며, 모집단을 대신하여 데이터를 획득하기 위한 대상
- ◆ 모집단으로부터 표본을 뽑는 과정을 추출(sampling)이라고 함
 - 표본은 모집단으로 가정하는 표본프레임(sample frame)으로부터 추출
 - ❖ 예: 국내 200개 4년제 대학별 학적DB가 표본프레임이 될 수 있음
 - 표본에 모집단의 특성이 잘 반영될 수 있도록 추출해야 표본에 대한 선택편의(selection bias)가 적고, 모집단에 대한 대표성이 높아짐
 - ❖ 예: 국내 200개 4년제 대학별로 재학생의 5%를 학년/성별/전공 등을 감안하여 임의로 추출
 - 표본 개수는 n 으로 표시함

모집단과 표본

- 추론통계(inferential statistics)란?

- ◆ 통계량(statistics)은 표본에 적용한 함수이고, 모수(parameter)는 모집단 특성을 반영한 값
- ◆ 추론통계는 통계량을 이용하여 모수를 추정(estimation)하거나 가설을 검정(test)
 - 모수 추정에 사용되는 통계량을 추정량(estimator)이라고 함



귀무가설과 대립가설

- 가설(hypothesis)이란?

- ◆ 모집단을 대상으로 모수와 관련된 평서문 형식의 문장으로 검정 대상이 됨
- ◆ 가설에는 귀무가설(null hypothesis)과 대립가설(alternative hypothesis)이 있음
- ◆ 귀무가설을 H_0 로, 대립가설을 H_1 혹은 H_a 로 표시

귀무가설과 대립가설

- 귀무가설(H_0)

- ◆ 현재 사실이라고 받아들이거나, 변수들 간에 관계가 없다는 내용으로 반증의 대상이 되며, 형태상 등호를 포함
- ◆ 예시1) H_0 : 20대 남성의 평균신장(μ)은 170cm(μ_0) 이하다. $\rightarrow H_0: \mu \leq 170\text{cm}$
 - 20대 남성의 신장은 어떤 분포를 띠게 되는 확률변수(random variable: RV)이며 X 로 표시
- ◆ 예시2) H_0 : 고객의 온라인 사이트 방문회수(X)는 구매금액(Y)과 아무 관련이 없다.
 - 변수들 간의 상관관계(상관분석)나 인과관계(회귀분석)를 확인해야 함

귀무가설과 대립가설

- 대립가설(H_a)

- ◆ 귀무가설을 부정함으로써 새롭게 입증하려는 내용으로 형태상 등호 불포함
- ◆ 예시1) H_a : 20대 남성의 평균신장(μ)은 170cm(μ_0) 보다 크다. $\rightarrow H_a: \mu > 170\text{cm}$
- ◆ 예시2) H_a : 고객의 온라인 사이트 방문회수가 증가하면 구매금액도 증가한다.
 - 두 변수 간에 양의 인과관계(positive causality)가 존재함
- ◆ 귀무가설과 대립가설은 서로 배타적(exclusive)이며 포괄적(exhausted)이어야 함
 - 두 가설 중 하나가 채택되면(accepted), 다른 하나는 채택되지 못함(not accepted)