

금융인을 위한

# 통계와 데이터 분석 입문

- 모집단과 표본,
- 통계적 추론 알아보기 •



# 학습 내용

- ① 모집단과 표본
- ② 모수와 통계량
- ③ 통계적 추론
- ④ 확률, 확률 변수의 기본 개념



# 모집단과 표본

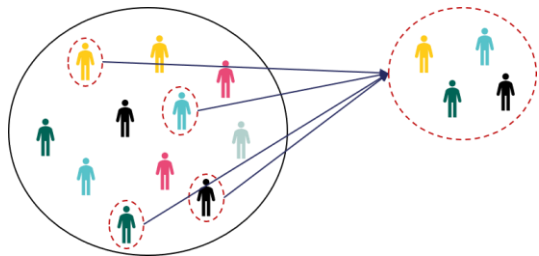
## 모집단(population)

우리가 정보를 얻고자 하는 전체 대상 또는 전체 집합

## 표본(sample)

모집단을 대표하기 위해 추출한 부분집합

→ 일반적으로 전체 집단을 알기 힘들거나 모든 자료를 수집하는 것은 비용·시간 제약으로 비효율적임



## 모집단과 표본

예

A은행 고객에 대해 현재 대출 상환능력을 여러 각도에서 살펴보기 위해 임의로 1000명의 고객을 조사함

- 모집단은 A은행 고객 전체
- 표본은 임의로 추출된 1000명의 고객

예

올해 10만대를 생산하려고 하는 자동차 회사에서  
안전성 테스트를 위해 50대의 차량에  
사람 모형 인형을 넣고 모의 충돌 실험을 시행함

- 이러한 파괴검사에서는 전수조사를 하는 것은 불가능함

## 모집단과 표본

- 표본을 통해 모집단의 특성을 정확하게 추정(estimation)하기 위해서는 표본을 추출할 때 허용할 수 있는 오차 또는 신뢰도를 만족할 수 있는 표본을 추출해야 함
  - ▶ garbage in, garbage out
    - 분석을 잘 해도 잘못된 자료로 분석한 경우 그 결과는 쓸모 없음
    - 신뢰성을 줄 수 있는 최소한의 표본의 크기를 찾고 적절한 표본추출 방법 활용해야 함

## 좋은 표본의 조건

- 관심있는 특성을 가진 개체들의 의식적 선택을 피해야 함

예

서비스 만족도 관련 조사하는 경우,  
만족도가 높거나 낮은 그룹을  
의도적으로 표본에 포함시키는 것(X)

## 좋은 표본의 조건

- 연구목적에 대한 결과를 얻기 위한 목적으로 표본을 선택하지 않도록 해야 함

예

주식시장을 분석하는 경우, 높은 수익률 또는 변동성을 보이는 주식을 선택하여 전체 집단으로 확대하여 해석하는 것(X)

- 표본추출 이론에 의하여 추출된 표본을 임의로 다른 것으로 대체하는 것은 주의해야 함
- 조사실시 중 추출된 표본의 조사 누락(무응답) 또는 거짓응답이 많으면 좋은 표본이 될 수 없음

## 표본추출 방법

- 모집단에서 무작위로 표본을 추출하여 선택하는 방법 중 가장 단순한 것은 제비뽑기식 임의추출 방법
- 모집단의 규모가 작을 때는 표본추출이 간단할 수 있음
- 규모가 커서 수십만, 수백만 또는 이보다 더 큰 경우 그 중에서 일부를 선택하는 것은 간단하지가 않음
  - ▶ 그 모집단을 잘 섞어서 뽑혀야 할 표본을 추출하는 것은 실제로는 상당히 어려움
- 모든 개체가 선택될 확률이 동일해야 함
- 특정 부분만이 표본으로 뽑힐 가능성을 제거하기 위해서 각 부분별로 나눠서 표본을 추출하는 방법도 고려해야 함



# 모수와 통계량

## 모수(parameter)

모집단의 특성을 나타내는 양적인 측도

## 통계량(statistic)

표본에서 얻어진 특성치들

- 모평균( $\mu$ )  $\Leftarrow$  표본평균( $\bar{X}$ )
- 모분산( $\sigma^2$ )  $\Leftarrow$  표본분산( $S^2$ )

## 통계적 추론(statistical Inference)

표본의 통계량을 이용하여 모집단의 모수에 대해서 추측하려고 하는 것

- 통계적 추론(statistical Inference)은 크게 추정(estimation)과 가설검정(hypothesis testing)으로 구성됨

## 추정

표본을 통해 모집단의 특성을 추측하는 과정

## 예

표본평균 계산을 통해 모집단 평균을 추측해보거나,  
모집단 평균에 대한 95% 신뢰구간을 계산해보는 것

- 통계적 추론(statistical Inference)은 크게 추정(estimation)과 가설검정(hypothesis testing)으로 구성됨

### 가설검정

모집단 모수에 대한 가설에 대해 표본이 가지고 있는 정보를 이용해 모집단 모수에 대한 가설이 올바른지 그렇지 않은지 판정하는 과정

- 모집단의 성질과 완벽하게 동일한 표본집단을 찾는다는 건 불가능함
- 아무리 잘 된 연구라도 통계량과 모수 사이에는 차이가 존재할 수밖에 없음
  - 표본오차(sampling error) 존재함

## 통계적 추론

- 통계적 추론을 위해서는 평균, 중위수 등의 통계량(statistic)을 확률분포를 가지는 확률변수라고 간주할 것
- 표본을 모집단으로부터 추출할 수 있는 많은 표본 중 하나로 간주
- 통계량의 표본분포(sampling distribution)에 대해서 알아야 함

예

표본평균의 표본분포

예

표본분산의 표본분포

# 확률, 확률 변수의 기본 개념

## 확률(probability)

- 불확실성의 수학적 표현
- 사건 A가 발생할 확률을  $P(A)$ , 사건 B가 발생할 확률을  $P(B)$ 라 할 때 아래의 성질을 만족함
  - ▶  $0 \leq P(A) \leq 1$
  - ▶ 사건 A가 일어나지 않을 확률  $P(A^c) = 1 - P(A)$
  - ▶ 사건 A, B에 대해서  $P(A \text{ or } B) = P(A \cup B)$   
 $= P(A) + P(B) - P(A \cap B)$
  - 상호 배반(disjoint)인 사건 A, B에 대해서  
 $P(A \cup B) = P(A) + P(B)$

## 확률, 확률 변수의 기본 개념

### 확률(probability)

- ▶ 서로 독립(independent)인 사건 A, B에 대해서

$$P(A \cap B) = P(A)P(B)$$

- ▶ 사건 B가 일어났을 때, 사건 A가 일어날 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- 독립인 사건 A, B이면  $P(A|B) = P(A)$

## 확률, 확률 변수의 기본 개념

### 확률변수(random variable)

모집단에 대해 어떤 구하고자 하는 것이 있을 때,  
어떤 값을 가질지에 대한 가능성을 확률로  
표현할 수 있음

예

한 기업이 투자적격 신용 등급에 속하는지 여부에  
관심 있다면 이를 확률변수(random variable)라고  
할 수 있음

- 모집단(population)은 우리가 정보를 얻고자 하는 전체 집합이고, 표본은 모집단을 대표해주는 모집단의 부분집합
- 표본을 통해 모집단의 특성을 정확하게 추정(estimation)하기 위해서는 표본을 추출할 때 허용할 수 있는 오차 또는 신뢰도를 만족할 수 있는 표본을 추출해야 함
- 모수(parameter)란 모집단의 특성을 나타내는 양적인 측도이고, 통계량(statistic)이란 표본에서 얻어진 특성치임
- 표본의 통계량을 이용하여 모집단의 모수에 대해서 추측하려고 하는 것을 통계적 추론(statistical inference)이라 하며 이를 위해서는 통계량의 표본분포를 이해해야 함