# 층화 표본 추출 (Stratified Sampling)

## 1. 층화 표본 추출의 정의

**층화 표본 추출**은 모집단을 서로 겹치지 않는 여러 개의 동질적인 그룹, 즉 ‘층(Strata)’으로 나눈 뒤, 각 층에서 독립적으로 단순 무작위 표본을 추출하는 방법입니다. 이렇게 각 층에서 추출된 표본들을 모두 결합하여 최종 표본을 구성합니다.

## 2. 사용하는 이유

* 모집단이 이질적인 여러 하위 집단으로 구성되어 있을 때, 각 집단의 특성을 정확하게 반영하고 전체 표본의 대표성을 높이기 위해 사용됩니다. 각 층이 모집단 내에서 차지하는 비율에 맞게 표본을 추출함으로써, 특정 집단이 과대 또는 과소 대표되는 것을 방지할 수 있습니다.
* 전체 모집단을 대상으로 하는 것보다 더 작은 오차 범위로 높은 정확도의 추정치를 얻을 수 있습니다.

## 3. 추출 과정

1. 1) 모집단을 층으로 나누기 (Stratification)

모집단의 특성을 나타내는 변수(예: 연령, 성별, 지역, 소득 수준 등)를 기준으로, 내부적으로는 동질적이고 외부적으로는 이질적인 여러 개의 층으로 분할합니다.

1. 2) 각 층의 표본 크기 결정 (Allocation)

각 층에서 몇 개의 표본을 추출할지 결정합니다. 주로 사용되는 방법은 다음과 같습니다.

* 비례 배분법: 각 층의 크기가 모집단에서 차지하는 비율에 따라 표본 크기를 할당합니다.
* 불비례 배분법: 각 층의 표준편차나 중요도 등을 고려하여 표본 크기를 할당합니다.

1. 3) 각 층에서 표본 추출 (Sampling)

각 층 내에서 단순 무작위 추출(Simple Random Sampling)과 같은 확률 표본 추출 방법을 사용하여 정해진 크기만큼의 표본을 독립적으로 추출합니다.

1. 4) 표본 결합

각 층에서 추출된 모든 표본을 하나로 합쳐 최종 표본을 구성합니다.

## 4. 장점 및 단점

**장점**

* 표본의 대표성 확보
* 추정의 정확도 향상
* 특정 하위 집단에 대한 분석 가능

**단점**

* 모집단에 대한 사전 정보(층화 변수)가 필요함
* 설계 및 실행 과정이 단순 무작위 추출보다 복잡함
* 층을 잘못 나누면 오히려 오차가 커질 수 있음

## 5. 예시

**상황:** 어떤 고등학교의 전체 학생 1,000명을 대상으로 만족도 조사를 하려고 합니다. 이 학교는 1학년 400명, 2학년 300명, 3학년 300명으로 구성되어 있습니다.

**추출 과정 (비례 배분법):**

1. 1) 층 나누기: 학생들을 1, 2, 3학년으로 층을 나눕니다.
2. 2) 표본 크기 결정: 전체 표본을 100명으로 정하고, 학년별 비율(4:3:3)에 따라 표본 크기를 할당합니다.

* 1학년 표본: 100명 \* (400/1000) = 40명
* 2학년 표본: 100명 \* (300/1000) = 30명
* 3학년 표본: 100명 \* (300/1000) = 30명

1. 3) 표본 추출: 각 학년 명단에서 무작위로 해당 인원수만큼 학생을 추출합니다.