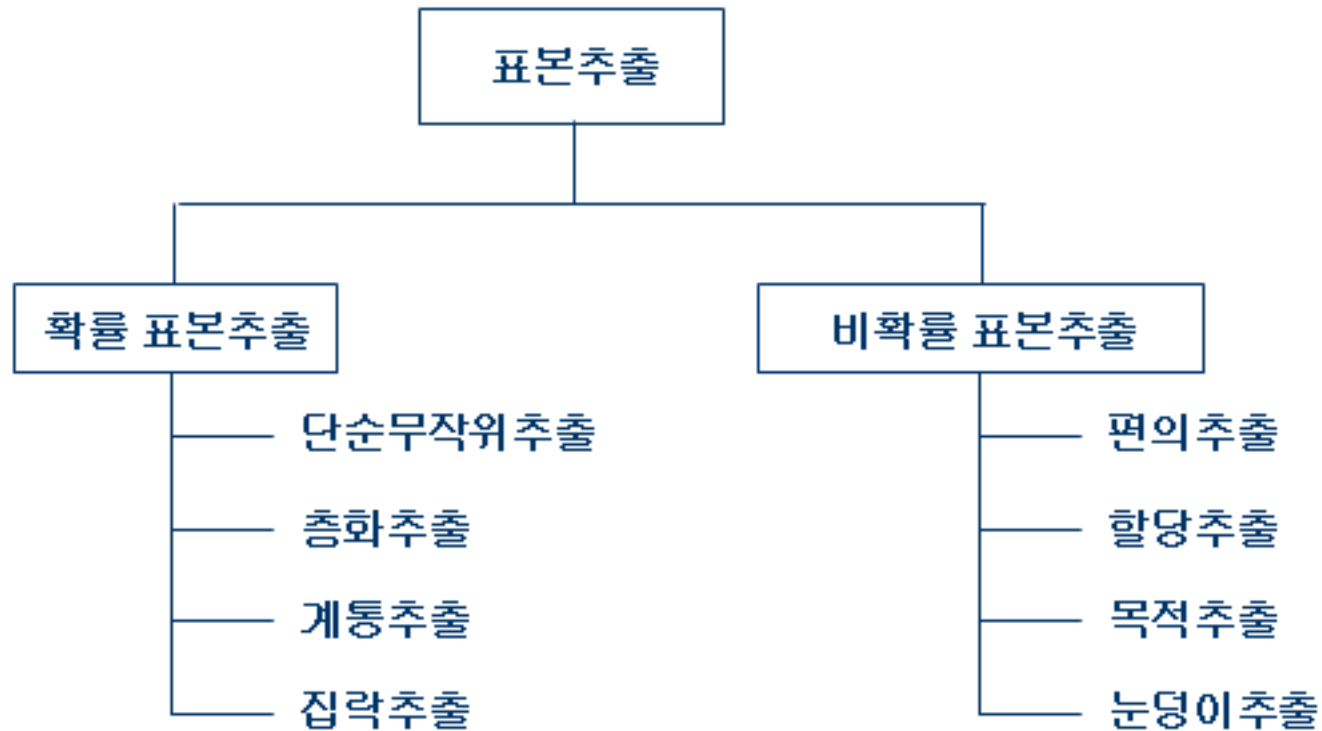


# 표본추출방법

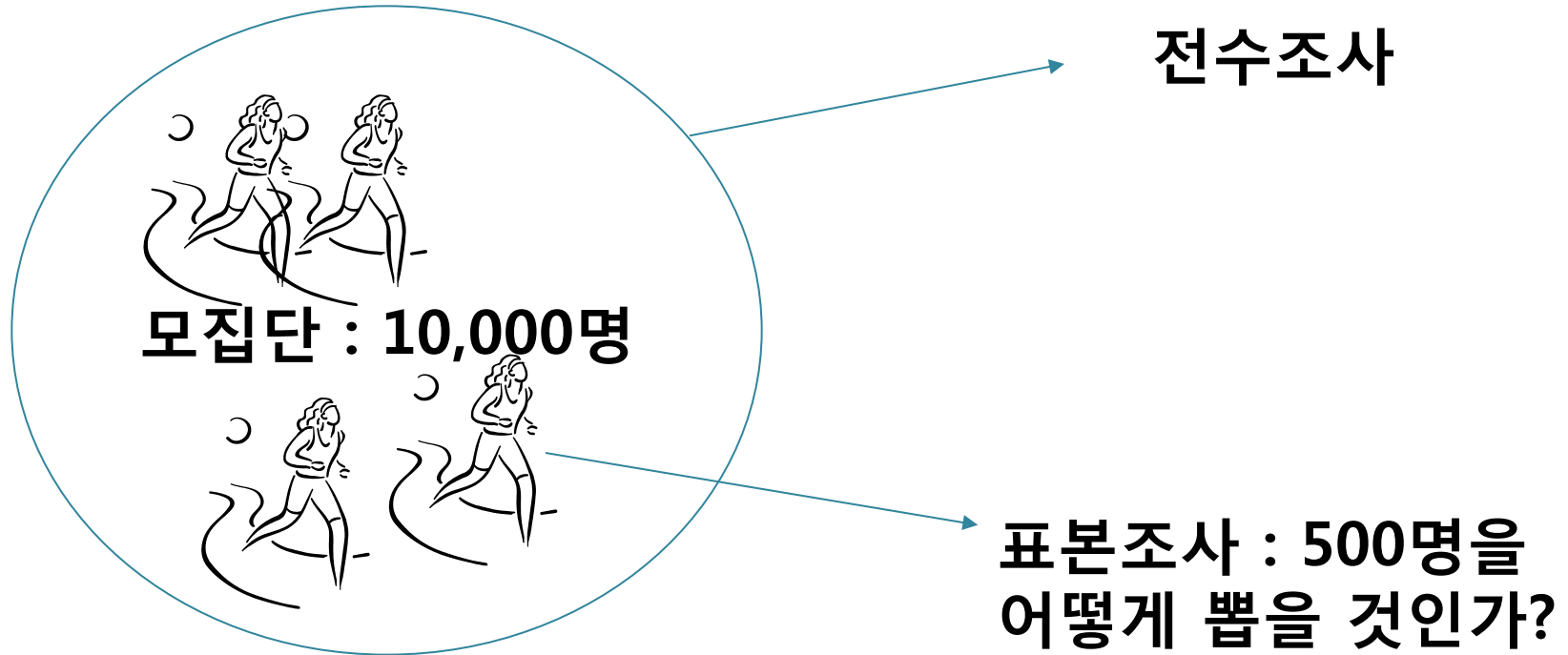


\* 실제 표본조사 - 다단계 집락표집법, 다단계 층화집락표집법

# 확률 표본추출 방법

1. 단순임의 추출법
2. 층화추출법
3. 집락추출법
4. 계통추출법

# 예) 원광대학생을 대상(500명)으로 건강조사 실시

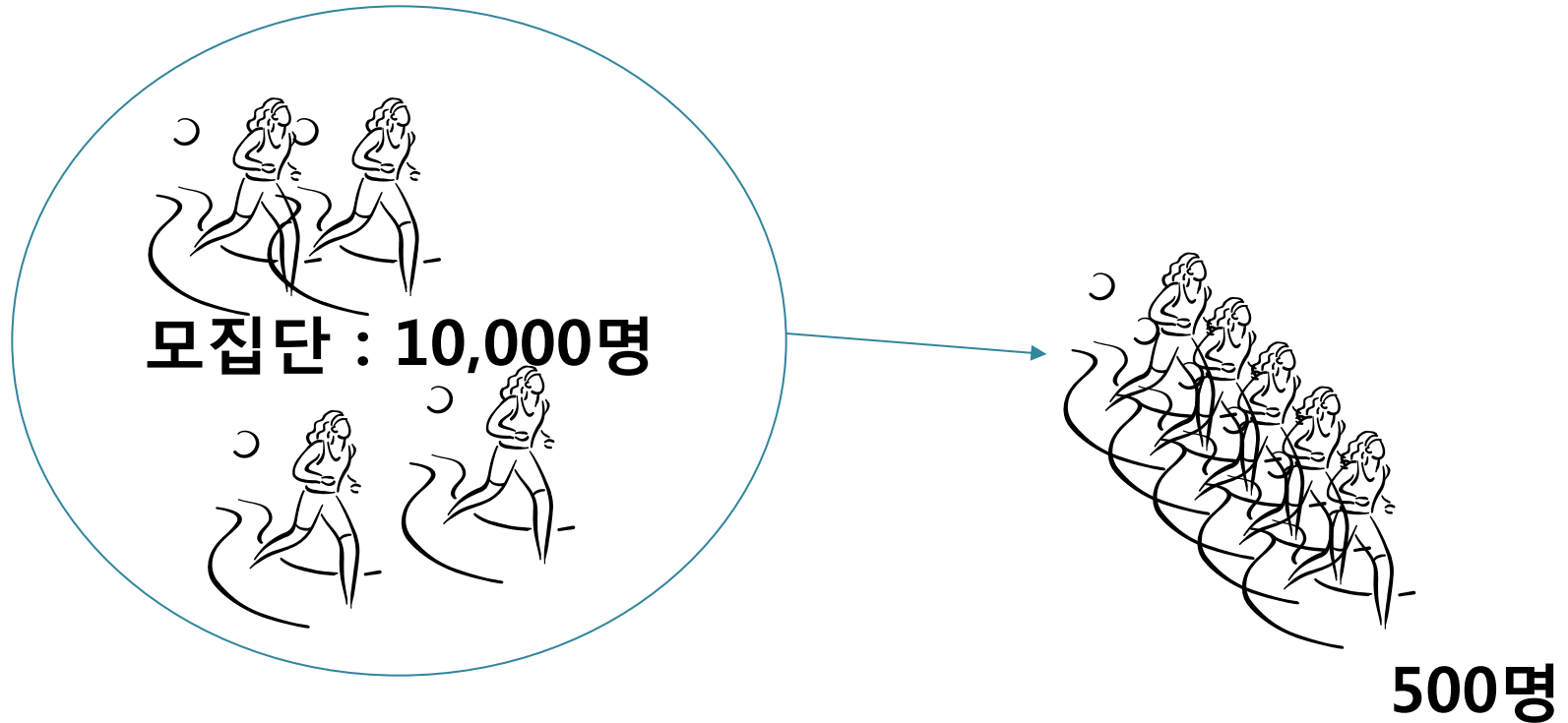


# 1. 단순임의 추출법 Simple Random Sampling(SRS )

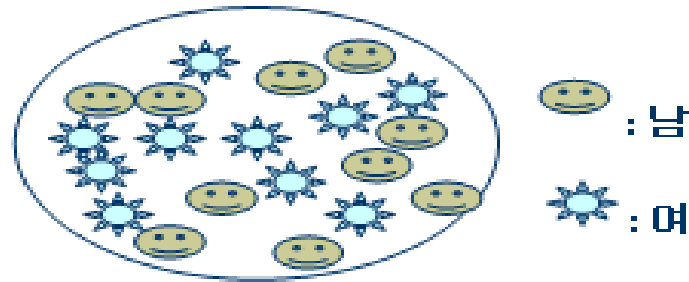
- 개념

모집단 전체의 일련번호를 부여해서 표본조사 틀을 만든 후, 난수표 등을 이용하여 각 개체가 뽑힐 가능성이 동일하게 되게끔 표본을 추출하는 방법.

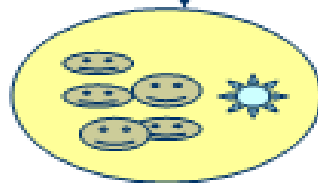
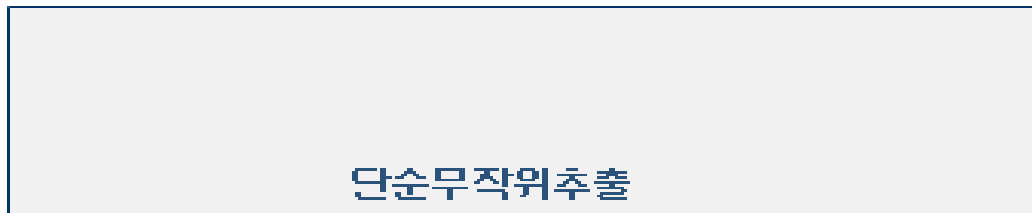
# 단순임의 추출법



# 단순임의 추출법



모집단 : 10,000명



500명

- ① 조사대상자 전체에 일련번호를 부여한다.
- ② 난수표 또는 컴퓨터를 이용하여 필요한 표본수 만큼 난수를 생성한다
- ③ 생성된 난수에 해당하는 일련번호를 가진 자를 표본으로 선정한다.

- 장점

1. 모집단에 대한 사전 지식이 불필요.
2. 추출기회가 동등하고 독립적이기 때문에 추출된 표본의 대표성이 높다.
3. 자료의 분류에 있어 오차의 개입이 적다.

- 단점

1. 모집단에 대해 가지고 있는 지식을 활용할 수 없다.
2. 비교적 표본의 규모가 커야 한다.
3. 표본프레임의 작성이 어렵다.

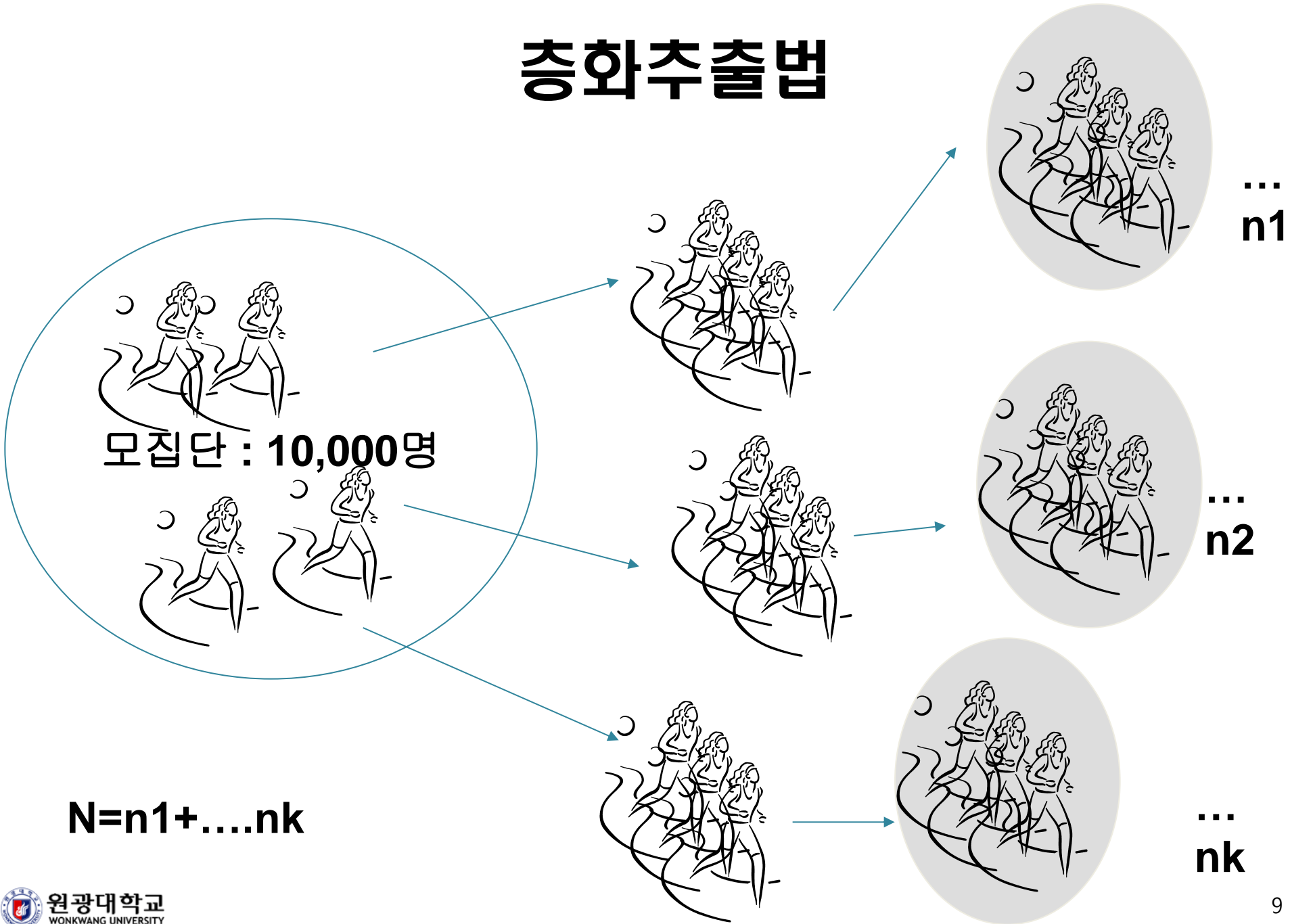
## 2. 층화추출법 Stratified Sampling

- 개념

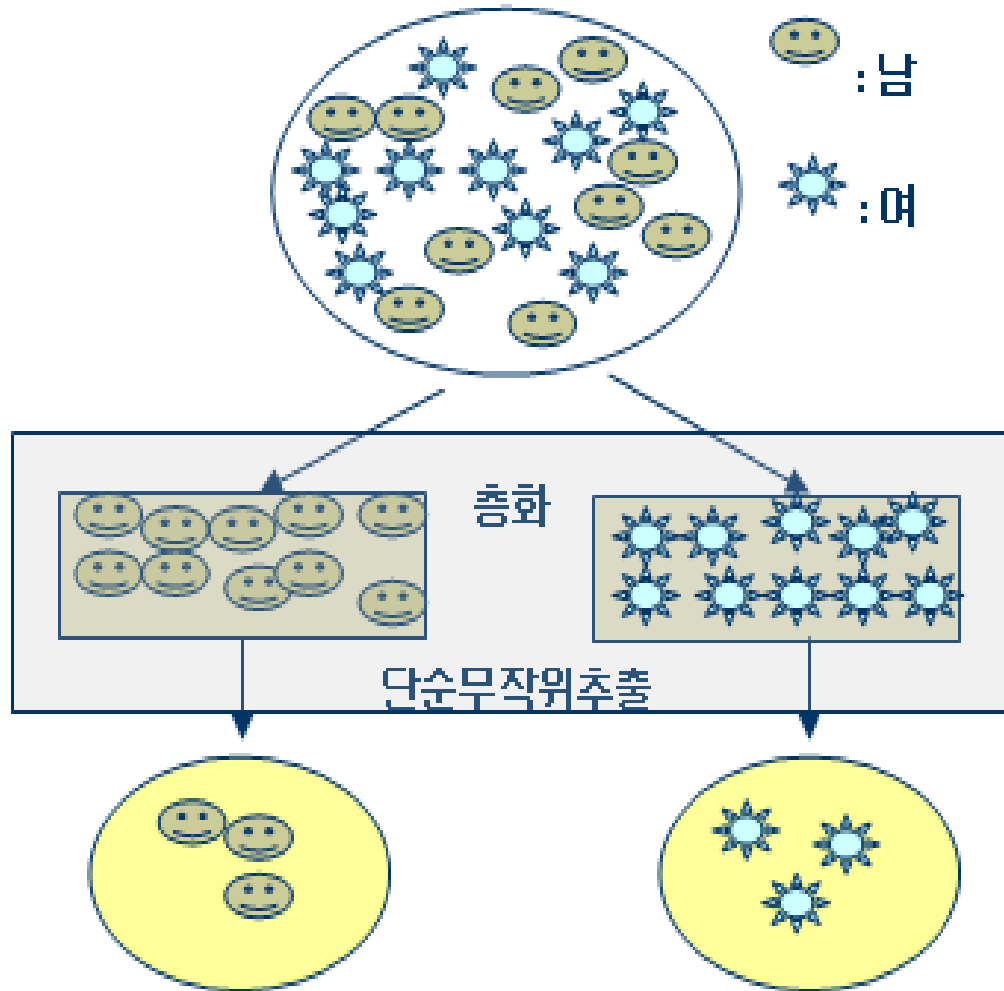
모집단을 먼저 서로 겹치지 않는 여러 개의 층으로 분할한 후, 각 층에서 단순임의추출법에 따라 배정된 표본을 추출하는 방법.



# 증화추출법

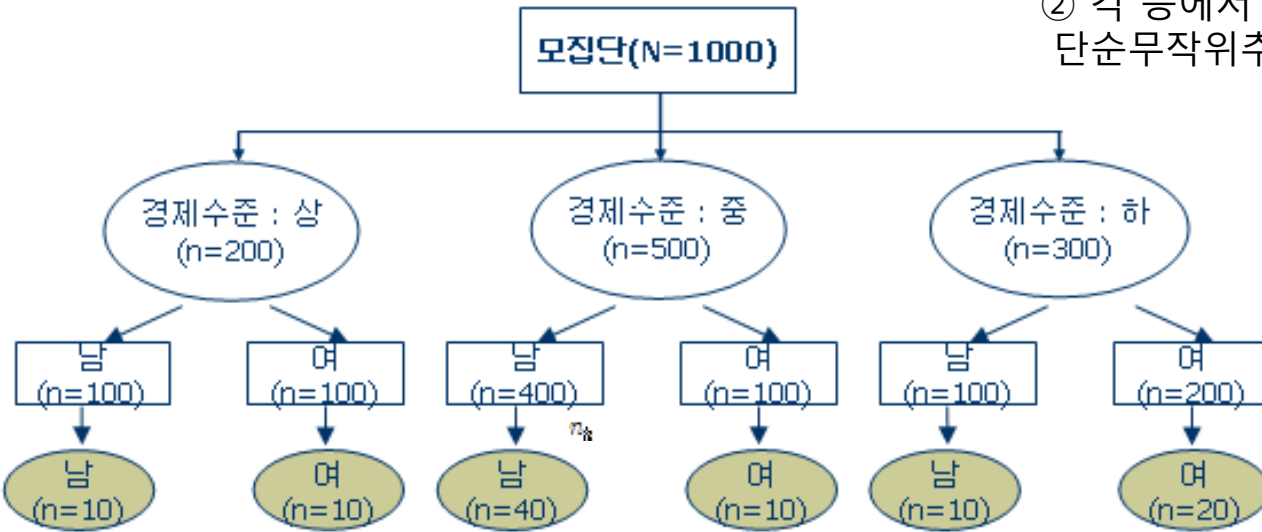


# 증화추출법

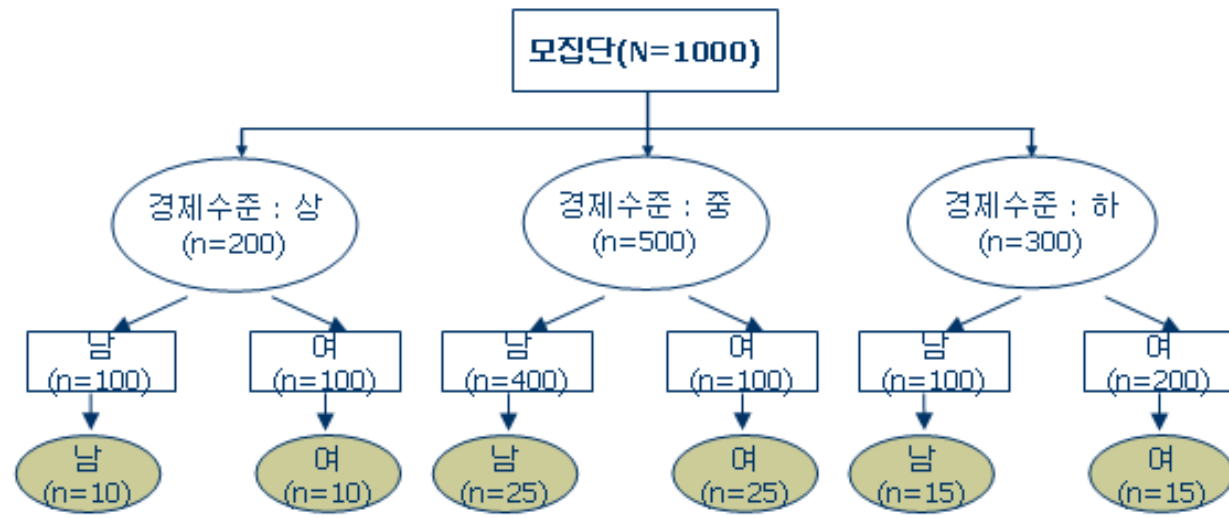


# 층화추출법

- ① 모집단의 L개의 층으로 층화한다.
- ② 각 층에서 표본크기를 결정하고, 단순무작위추출법에 따라 표본을 추출한다.



비례층화추출



비비례층화추출

- 모집단을 특성에 의해 몇 개의 그룹(strata)으로 나눈다.
- 그룹 간에는 차이가 존재한다. 그러므로 각 strata에서 골고루 개체를 선택해야 한다.
- 각 그룹에서 SRS방법에 의해 일정 수만큼 개체를 임의로 선택한다.
- 일정 수? (그룹에 속한 개체 수)모집단 개체수 비율  
 (예) 학생들의 IQ조사 위해 500명 추출한다면?
  - 원광대학교: (10,000)=경상대(5,000)+이과대(2,000)+공과대(3,000)
  - 단과대학별로 그룹(strata)화
    - 경상대:  $500 \times 0.5 = 250$ 명/이과대=100명/공과대=150명
- Multi-stage 층화 추출
  - 전과 동일 =>그룹을 여러 단계로 나눔(대학별, 학년별, 성별)

- 장점

1. 중요 집단은 빼놓지 않고 표본에 포함시킬 수 있음.
2. 동질적 대상은 표본의 수가 적어도 대표성이 높음.
3. 각 층의 특성에 대한 추정과 비교가 가능.

- 단점

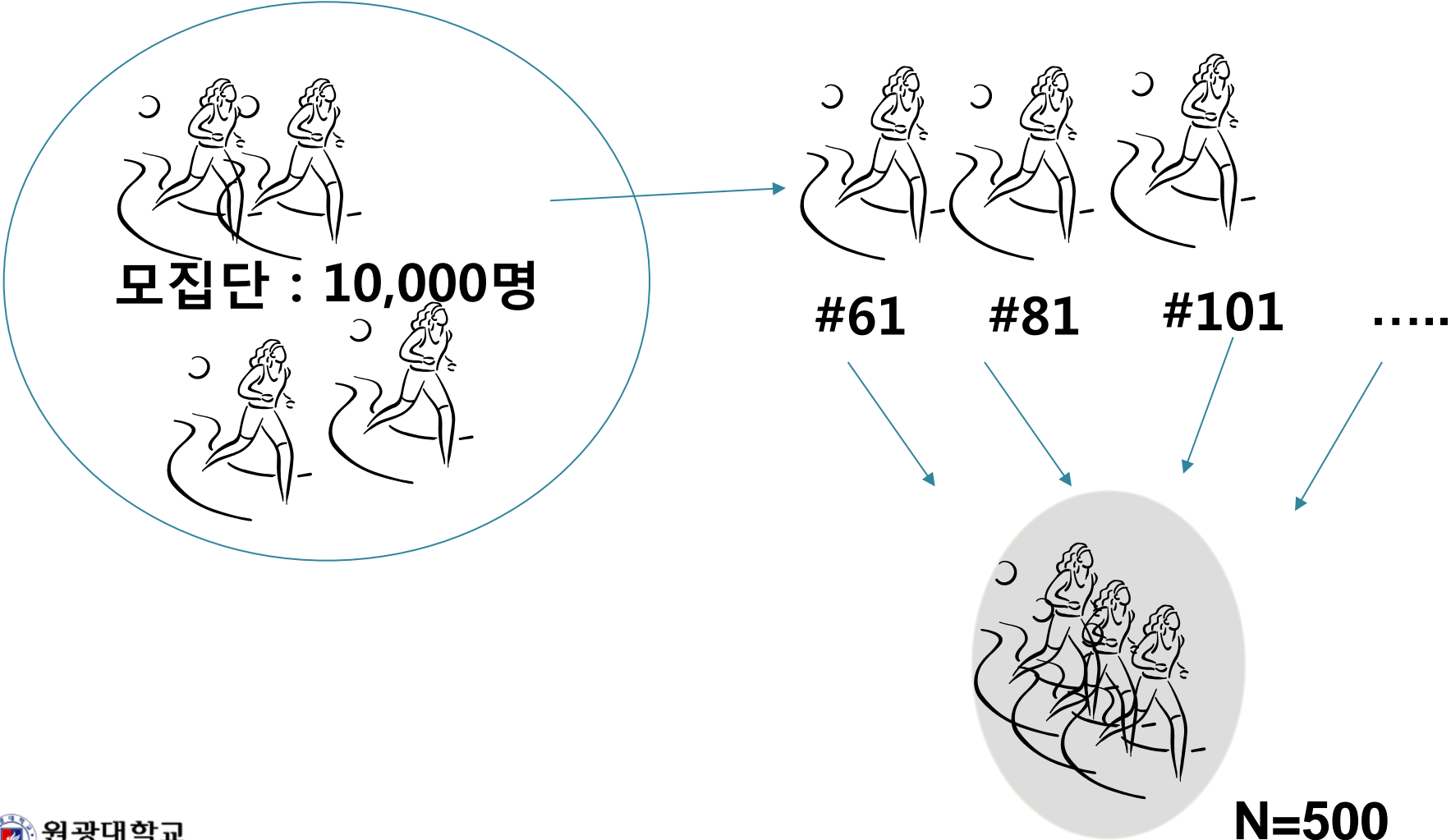
1. 원형으로 복귀하기 어렵다.
2. 층화시 시간과 노력이 소요.
3. 모집단에 대한 지식이 필요.

### 3. 계통추출법 (Systematic Sampling)

- 개념

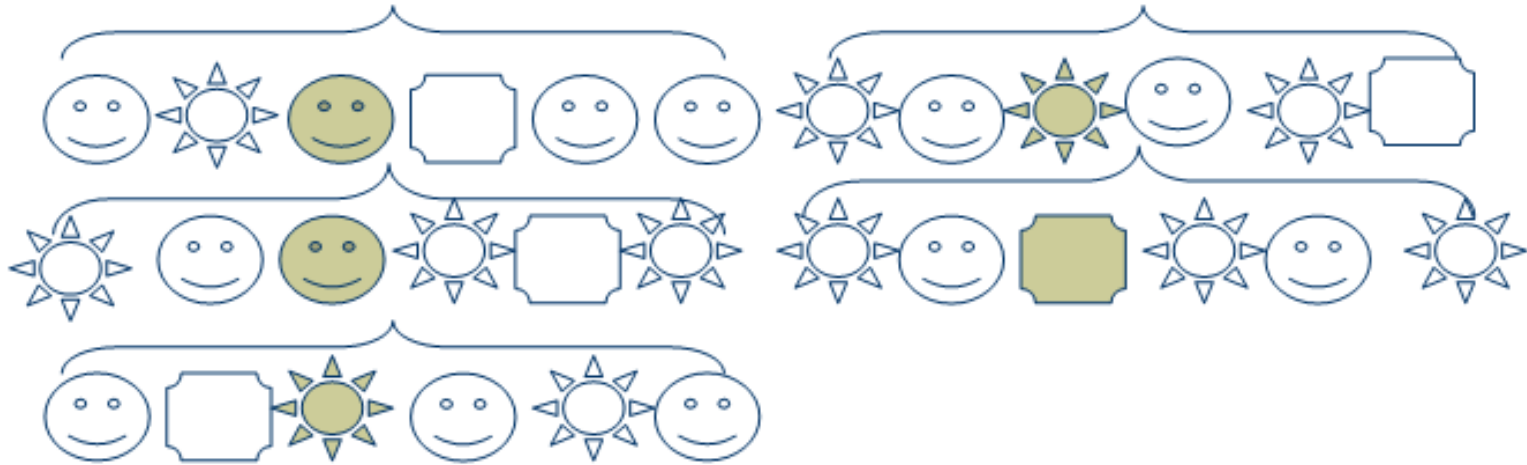
추출 틀에서 처음  $k$ 개 단위들 중에서 랜덤하게 하나의 단위를 추출하고 그 이후 매  $k$ 번째 간격마다 하나씩의 단위를 표본으로 추출하는 표본추출방법.

# 계통추출



# 계통추출법

- ① 추출틀에서 처음  $k$ 개 단위 중에서 랜덤하게 하나  $r(1 \leq r \leq k)$ 를 선택한다. 이때  $r$ 을 출발점(starting point)이라고 한다.
- ② 이후  $r, r+k, r+2k, \dots, r+(n-1)k$  들을 표본으로 추출한다.



30명으로 구성된 모집단에서 5명을 뽑고자 하는 경우 추출간격  $K$ 는 6이 된다. 30명을 6명씩 묶고, 첫번째 6명에서 단순무작위추출에 의해 1명을 뽑고, 같은 순서에 배치된(이번 예에서는 세번째) 대상자를 차례로 뽑아간다.



- SRS방법과 같이 모집단의 개체에 일련번호를 부여한다.
- 1번부터(모집단 수/표본 수)번까지 개체 중 난수를 이용하여 하나를 선택한다.
- 그 개체에 부여된 일련번호에 일정 수만큼( $k$ ) 더해 그 번호의 개체를 선택한다
- 단순임의추출법과 유사
- 모집단을 잘 대표하는 표본을 추출할 수 있고 SRS에 비해 간편하다.
- SRS와 동일한 문제점 모집단이 큰 경우 일련번호 부여에 문제가 있다.

- 장점

1. 표본추출이 간편.
2. 단순임의추출법의 대용으로 사용.
3. 일반적으로 표본이 모집단 전체를 잘 반영.

- 단점

1. 모집단 추출틀에 대한 추가적인 가정이 필요.
2. 표본의 대표성이 저해

(주기성, 특정경향성을 보일 땐 피한다.)

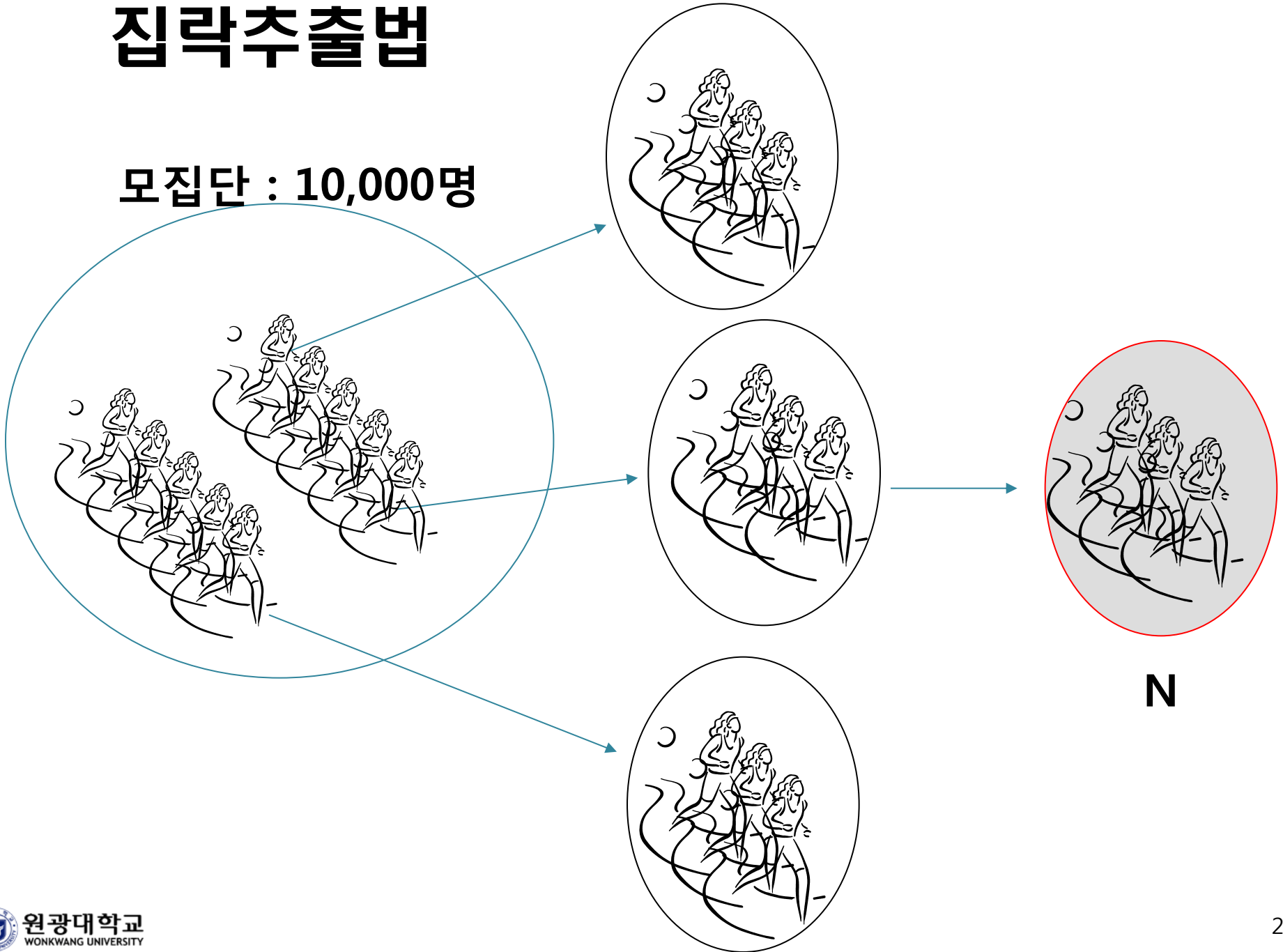
## 4. 집락추출법 (Clustering Sampling)

- 개념

모집단을 서로 인접한 기본단위들을 묶어 집락을 구성하여 먼저 집락을 추출하고 추출된 집락 내의 일부 또는 전체를 조사하는 방법.

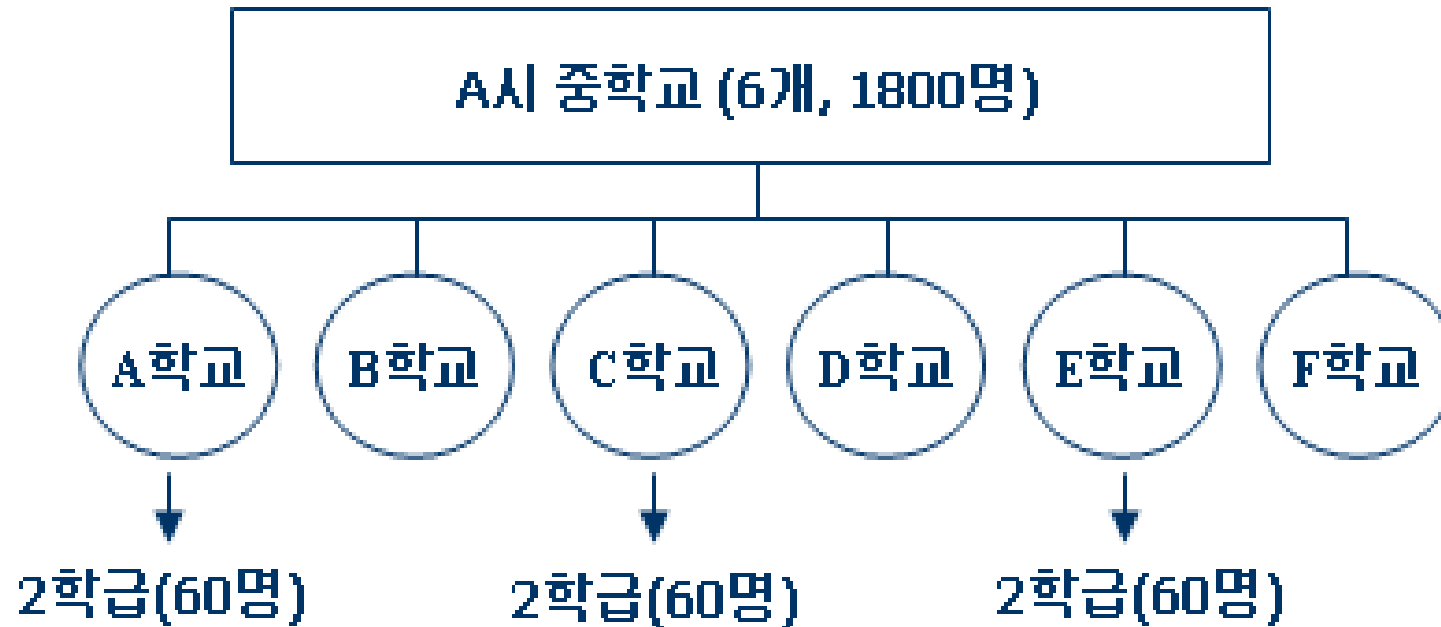
# 집락추출법

모집단 : 10,000명



N

# 집락추출법



- 모집단을 특성에 의해 몇 개의 그룹(cluster)으로 나누고 그룹에 번호를 부여한다.
- 반드시 그룹 간에는 차이가 없어야 한다
  - 차이가 있다면 조사에 오류
- SRS방법과 같이 난수를 이용하여 그룹 중 하나를 임의로 선택한다.
- 선택된 그룹에 있는 개체를 모두 선택한다.
- 원하는 표본 수를 얻을 때까지 계속한다.
- 마지막 그룹처리 문제: 남은 표본 수가 500개인데 선택된 그룹 안에 600명이 있다면
  - 500명만 단순 임의 추출하면 된다.
  - 표본 선택 시간과 비용을 줄일 수 있다.
  - 그룹 간에 차이가 있다면 조사 결과가 왜곡될 수 있다.

- 장점

1. 모집단의 목록이 없을 때 이용.
2. 현지조사 비용을 절감.
3. 각 군집의 속성으로 모집단의 속성을 추측.

- 단점

1. 각 군집의 이질성을 확보하기 위한 분류의 기준 설정이 어려움.
2. 따라서, 다른 확률표본추출에 비해 표본오차가 더 커질 가능성이 있다.

# 표본추출방법 정리

- 표본추출방법에는 여러 방법들이 있다.
- 표본추출론에 관한 어느 교과서들에서도 기본적으로
  - 단순랜덤추출(단순무작위 추출; simple random sampling)  
또는 계통추출(systematic sampling)
  - 층화추출(stratified sampling)
  - 집락추출(cluster sampling) 등의 서너 방법을 다루고 있는데,
- 실제로는 이들 방법들이 개별적으로 사용되는 것이 아니라 서로 결합된 방식으로 사용된다.
  - 대표적으로 층화 집락 무작위추출(stratified cluster random sampling) 이 쓰이고,
  - 이 때 집락추출의 단계수에 따라 그것이 3단이면 층화 3단 집락추출(stratified three stage cluster sampling)라고 한다.



# 예제 1

어느 사립대학에서 학생들은 다음과 같은 체계에 따라 구별되어졌다.

분류	학생수
4학년	150
3학년	163
2학년	195
1학년	220

여기서, 우리가  $n=40$  크기의 층화확률표본을  
추출하기 위하여 비례배분을 사용한다면, 각 층  
으로부터 우리가 추출해야 할 표본의 크기는 얼  
마인가?

# -> 풀이

$n=40, N_1=150, N_2=163, N_3=195, N_4=220,$   
 $N=728$  이다.

$$n_1=(150/728)40=8, \quad n_2=(163/728)40=9,$$
$$n_3=(195/728)40=11, \quad n_4=(220/728)40=12$$

따라서,

분류	학생수
4학년	8
3학년	9
2학년	11
1학년	12

# 비확률 표본추출 방법

- 편의표본추출(*convenience sampling* ; 대표적) : 조사자나 면접원이 편리한 장소와 시간에 접촉하기 편리한 대상들을 표본으로 추출하는 것으로 표적모 집단 구성원들은 동질적(*homogeneous*)이어서 어떤 구성원을 대상으로 조사하더라도 마찬가지로 하는 것을 가정.
  - 장점 : 조사대상을 적은 비용과 시간으로 확보할 수 있으며, *construct* 개발, 설문지 개발 후 사전조사(*pretest*) 등에 사용될 수 있고 편의표본으로부터 엄격한 분석결과를 획득하지는 못하지만 조사대상들의 특성에 대한 개괄적 정보를 획득할 수 있다.
  - 한계점 : 표본의 모집단 대표성이 매우 낮다.
- 할당표본추출(*quota sampling*) : 인구통계적 특성(나이, 성별, 소득수준 등), 거주지 등의 측면에서 사전에 정해진 비율에 따라 모집단 구성원들을 할당하는 방법(← 층화표본추출과 구별필요).
  - 층화표본추출 : 조사하고자 하는 특성 면에서 모집이 여러 가지 다른 집단들로 구성된 경우 각 집단에서 일부씩 추출하는 방식(사전에 그룹화 가능).
  - 할당표본추출 : 인구통계적 특성이나 거주지를 중심으로 조사자가 표본의 크기를 할당(사전에 그룹화 정보없이 모집단을 구분).
  - 장점 : 모집단에 대한 대표성이 단순무작위표본에 비해 높다.
  - 한계점 : 할당표본추출을 하기 위해서 모집단의 특성에 대한 사전지식에 필요하며 이를 전혀 모르는 경우 할당표본추출은 불가능.

- 판단표본추출(*judgment sampling*) : 조사자가 조사목적에 적합하다고 판단되는 구성원들을 표본으로 추출하는 것으로 해당분야의 전문가들의 의견이 표적모집단의 대표성을 갖는다고 가정.
  - 장점 : 해당분야의 전문가로 판단되어 선정된 표본이 실제로 유용한 정보를 제공할 수 있다면 판단표본추출법은 매우 유용.
  - 한계점 : 판단표본이 편의표본보다 더 대표성을 가질 것이라는 것은 가정이나 기대일 뿐, 실제로 모집단의 대표성 정도는 평가할 수 없다.
- 눈덩이 표본추출 (*snowball sampling*) : 조사자가 적절하다고 판단되는 조사대상자들을 선정한 다음 그들로 하여금 또 다른 조사대상자들을 추천하도록 하는 방법으로 처음에는 조사자의 판단에 따라 조사대상자들이 선정되므로 판단표본추출법의 일종이라 할 수 있다.
  - 장점 : 조사자가 표적모집단 구성원들 중 극소수 이외에는 누가 표본으로 적절한지를 판단할 수 없는 경우 사용될 수 있다(*FGI*같은 비계량적 조사에서 흔히 사용).
  - 한계점 : 연속적 추천에 의해 선정된 조사대상자들에는 동질성이 높을 수 있으나 표적모집단과는 매우 유리된 특성을 가질 수 있다.

# 표본추출법 결정시 고려할 요인들

- 가장 효율적인, 적합한 추출법을 선택할 것
  - 조사의 목적 : 표적모집단의 특성을 정확히 추정하는 것이 매우 중요하다면 표본의 모집단 대표성이 높도록 확률표본추출을 해야 하나 표적모집단의 특성에 대한 개괄적인 아이디어를 얻고자 한다면 편의표본을 이용할 수도 있다.
  - 표적모집단에 대한 사전지식 : 체계적 표본추출, 층화표본추출, 할당표본추출을 하기 위해서는 표적모집단에 대한 사전지식이 필요하며, 판단표본과 눈덩이 표본추출의 경우도 표적모집단에 대한 사전지식이 요구됨
    - (사전지식이 전무하다면 단순무작위 혹은 편의표본추출을 할 수밖에 없음)
- 조사의 용이성, 조사비용 등을 함께 고려할 것
  - 가능한 한 단순한 추출법을 사용할 것
  - 시간 : 시간적 제약이 큰 경우 비확률표본추출을 함으로써 시간절약을 할 수 있다.
  - 예산 : 예산이 한정적인 경우 정확한 정보획득을 포기하더라도 비확률표본추출을 할 수 밖에 없다.