

## “확률과 통계” 참고 자료

■ **표본추출 방법** : 표본추출을 통한 통계적 방법을 사용하기 위해서는 추출과정에서 반드시 랜덤성의 조건이 들어가야 함

(1) **비확률추출**: 확률표본 추출보다 더 정확하거나 경제적인 수 있으나(모집단의 성질을 잘 알고 있는 경우) 오차에 대한 확률적인 분석이 불가능하므로 확률에 입각한 추정, 가설검정 등을 할 수는 없음. 따라서 일반적인 경우에는 피하는 것이 좋음

가. **판단추출법(judgement sampling)**: 전문가가 그의 생각하기에 가장 효율적이라고 생각되는 표본을 주관적인 판단 하에 추출함. 경우에 따라서는 확률표본추출보다 더 정확한 추정치를 구할 수도 있으나, 오차의 확률적 분석이 불가능함

나. **편의표본추출(convenience sampling)**: 연구자가 가장 손쉽게 얻을 수 있는 표본을 추출하는 방법으로 가장 경제적인 방법이나 표본오차가 클 것임

(2) **확률추출**: 모집단으로부터 표본으로 뽑힐 각 항목들의 선택이, 알려진 확률에 따라 이루어지는 표본추출방법, 오차가 확률로 예측될 수 있는 것이 특징임(확률적 표본-> 각 추출단위가 표본 가운데 포함될 확률이 얼마인가를 알 수 있음)  
예) 랜덤추출, 계통추출, 층화추출, 군집추출

가. **단순임의(랜덤) 추출법**: 가장 간단하고 다른 방법의 이론적 근거가 됨

- N개의 추출 단위가 있는 모집단에서 n개의 표본 단위를 비복원 추출할 때, (N n)개의 모든 가능한 경우가 동일한 확률로 추출되는 방법, 난수표 이용
- 일반적으로 복원추출보다 비복원추출이 더 바람직함

나. **층화임의(랜덤)추출법**: 모집단에 대한 어떤 사전 지식을 갖고 있을 때는 이를 표본추출에 이용하는 것이 바람직함. 그 한 방법으로서 층화랜덤추출이 있는데 이 방법은 모집단을 동질적이며 상호 배타적인 몇 개의 하위집단 또는 층으로 나눈 다음, 각 층에서 어떤 규칙에 의해서 표본추출을 하는 것임 (남녀별, 연령층별)

- 즉, 모집단을 동질적인 층으로 나눌 수 있을 때, 층화추출법을 사용하면 추정량의 분산을 줄일 수 있는 **장점**이 있음
- 층을 나누는 기본원리는 층간의 변이성(變異性)을 크게 하고 층 내부에서의 변이성을 작게 하는 것임-> 따라서 각 층에서는 상대적으로 작은 표본을 갖고도 정확한 추정을 할 수 있게 됨
- 각 층에 표본 크기를 할당하는 방법에는 각 층의 크기에 비례해서 각 층의 표본 크기를 정하는 비례할당방식과 추정량(X의 표본평균)의 분산을 최소로 하도록 표본 크기를 할당하는 최적할당 방식이 있음
- 층화추출이 가능할 때 **장점**
  - 1) 추정량의 분산이 줄어들어 더 적은 수의 표본으로도 정확한 추정이 가능하고
  - 2) 각 층별로 충분한 표본 수를 확보할 수 있으며,

- 3) 부분 집단별로 정도(精度)가 요구될 때 이에 응할 수 있고
- 4) 모집단의 분포가 한쪽으로 치우쳐 있다고 생각될 때 이에 대응할 수 있으며
- 5) 기타 행정상의 편의(편리한 점)가 있을 수 있음
- 적용 예: 한 도시의 월 생계비 조사 (고소득층 동쪽에 많이 살고 저소득층 서쪽에 많이 산다면)

#### 다. 군집(집락)추출(cluster sampling)

- 모집단을 군집 또는 집락(cluster)이라고 하는 부분 집단으로 나눈 다음, 몇 개의 표본 군집을 임의 추출하고 추출된 군집 내의 조사단위를 전수조사 하는 것을 말함
- 층화추출과는 반대로 군집추출에서는 각 부분집단들이 모집단과 같은 특성을 갖고 있을수록 좋음. 즉, 부분집단들 사이에 이질성이 적어야만 선택된 군집에서 모집단의 특성을 올바르게 추정해 낼 수 있음
- 추출된 각 군집 내에서 다시 추출단위를 선택하는 방법을 다단추출법이라 함.  
예컨대, 서울에서 가구당 월 생계비를 조사한다면 우선 동을 추출단위로 하여 전체 350개 동에서 표본 동을 랜덤추출한 다음 각 동에서 다시 표본가구를 임의추출 함. 그러면 이 추출법을 2단 군집추출법이라고 하며, 동을 1차 추출단위라 하며 가구를 2차 추출단위라 함
- 모집단의 크기가 작으면 단순랜덤추출법이나 층화추출법이 가능하지만, 모집단의 크기가 크면 단순랜덤추출은 몇 가지 어려움이 있음(리스트 작성이 어려움, 조사 단위가 흩어져 있으므로 비용이 많이 듦, 조사 관리가 어려움)  
\* 리스트 작성이 어렵다는 것은 표본프레임이 없는 경우(교재 참고)에 해당
- 군집추출을 하는 목적 : 비용의 절약, 표본 관리의 용이(관리를 쉽게 하기 위한 것)

#### 라. 계통추출(systematic sampling)

- 표본을 시간적으로나 공간적으로 일정한 간격을 두고 취하는 추출법으로, 모집단과 표본의 크기를 각각 N과 n이라 할 때  $k=N/n$ 을 추출간격이라 하는데, 계통추출은 모집단의 처음의 k개의 추출단위에서 임의로 하나를 추출하고, 다음부터는 매 k번째마다 표본을 추출해 나가는 것임
- 예컨대, 1부터 10000까지 만 개의 단위가 있는 모집단에서 500개의 단위를 계통 추출하려면  $k=10000/500=20$ 이므로 처음 1번부터 20번 사이에서 임의로 하나를 추출하고 예컨대 17번이 택해졌다면 다음엔 37, 57, 77, .....이 취해짐, 또는 가나다 순으로 정리된 전화번호부나 주민등록표에서 표본추출을 할 때 계통추출법을 사용하면 매우 편리한 장점이 있음
- **장점**
  - 1) 방법이 간단함, 모집단 전체에서 골고루 표본이 추출됨
  - 2) 모집단에 장기적인(상향적인) 경향이 있으며 이를 사전에 알 수만 있다면 층화 추출과 비슷한 효과를 얻도록 계획을 세울 수 있음. 그러나 모집단의 배열순서에 예측하지 못했던 주기성이 있다면 계통추출에 의한 결과는 상당히 편의된 결과가 될 수도 있음 즉, 편의가 생기는 **단점**이 있음