

조사 및 통계분석

III. 표본추출

1. 확률 표본추출
2. 비확률 표본추출
3. 단순무작위 표본추출
4. 난수표를 이용한 SRS
5. 집락추출과 층화추출
6. 편의 표본추출

확률 표본추출

Probability sampling vs. Non-probability sampling



확률 표본추출

Probability sampling vs. Non-probability sampling

표본추출은 확률 표본추출과 비확률 표본추출로 구분한다.

확률 표본추출 (probability sampling)은 모집단의 모든 구성원들에 대하여 표본으로 뽑힐 수 있는 확률이 동일하게 부여된 표본추출방법이다. 반대로 **비확률 표본추출 (non-probability sampling)**은 각 구성원들이 표본으로 추출될 수 있는 확률이 동일하지 않아 표본의 특성이 모집단의 특성을 정확히 대표하지 못하고 어느 한 쪽으로 치우치게 (biased)되는 경우가 발생할 수 있다.

확률 표본추출은 모집단 구성원 중 표본으로 선정되는 대상 또는 대상으로 구성된 집단인 표본추출 단위 (sampling unit)에 대한 모든 목록, 즉 **표본 추출 프레임 (sampling frame)**을 이용해 동일한 확률로 표본 대상자를 추출한다. 이 경우, 추출된 표본은 속성 측면에서 모집단과 매우 유사하다고 할 수 있다.

편의 표본추출

Biased sampling

표본 통계량(a statistic)과 모수(the parameter)와의 차이가 어느 한 쪽으로 치우치는 것을 **편향(bias)**라 부른다. 표본추출에 있어 편향을 최소화하는 것이 매우 중요하다. 따라서 표본추출 시 가능한 한 편향을 가져다 줄 수 있는 개인적 선택의 기회를 배제하는 것이 매우 중요하다.

편의추출(convenience sampling)은 조사자가 자신이 접촉하기 쉬운 표본을 임의로 선택하는 방법이며, **자발적 응답(voluntary response sample)**은 응답자가 표본으로 참여할 것인가를 자신이 임의로 결정하는 방법이다. 결국 이 두 방법은 표본추출에 있어 조사자 또는 응답자가 표본 선택에 있어 개인의 선호가 작용하므로 조사 결과의 편향을 가져다 주게 된다.

다시 말해서 편의추출은 조사자가 표본을 선택하는 것을, 자발적 응답은 응답자가 표본으로 선택되는 것을 결정하게 되므로 모집단의 개체들이 표본으로 선택될 확률이 다르게 된다.

집락추출과 층화추출

Cluster sampling vs. Stratified sampling

모집단이 너무 커서 표본추출 프레임(sampling frame)을 작성하기 어려울 경우 모집단을 몇 개의 소그룹으로 나눈다. 그리고 이들 그룹의 리스트를 표본추출 프레임으로 하여 다수의 표본 그룹을 추출한 뒤, 이들에 대한 표본통계량을 산출하고, 이들 통계량을 논리적인 방법으로 조정하여 새로운 통계량을 만드는 방법을 사용하여 확률표본추출과 유사한 결과를 얻기도 한다.

이때 **집락추출법(cluster sampling)**과 **층화추출법(stratified sampling)**을 사용한다.

집락(cluster)과 층(strata)은 모집단을 구성하고 있는 소그룹이다. 그러나 **층**은 그룹 내 개체들이 동질적 특성을 갖고 있으며, **집락**은 그룹을 구성하고 있는 개체들이 이질적 특성을 갖는다는 점에서 다르다.

결국 집락은 모집단과 유사한 특성을 갖고 있는 소그룹이며, 층은 모집단을 구성하고 있는 개체들 중 각기 독특한 특성을 지닌 개체들로 구성된 소그룹이라고 할 수 있다.

집락추출과 층화추출

Cluster sampling vs. Stratified sampling



(그림) 표본 추출 방법

층화추출법(stratified sampling)

모집단을 각기 다른 특성을 가진, 그러나 구성원은 서로 유사한 성질을 갖는 2개 이상의 층(stratum)으로 구분하고, 각 층으로부터 단순무작위추출법을 적용하여 표본을 추출하는 방법.

- 1) 표본크기가 크지 않아도 모집단의 대표성이 보장
- 2) 단순무작위추출보다 자료의 분산을 축소
- 3) 각 층화 집단의 특성을 구별하여 알 수 있으며, 비교가 가능
- 4) 층 내의 개체들은 동질적이고, 층 간에는 이질적
- 5) 모집단의 각 층에 대한 정확한 정보 필요.

집락추출과 층화추출

Cluster sampling vs. Stratified sampling



(그림) 표본 추출 방법

집락추출법(cluster sampling)

모집단을 조사 단위 또는 집계 단위로 모아 군집들(clusters)로 나누고, 이들 군집들 중 일부 군집을 추출한 뒤 추출된 군집의구성원 일부 또는 전체를 표본으로 추출하는 방법

- 1) 군집 내의 개체들은 이질적이고, 군집 간(A,B,C) 동질적
- 2) 시간과 비용이 절약
- 3) 전체 모집단의 표본추출 프레임(sampling frame) 불필요
- 4) 단순 무작위 추출보다 측정 집단을 과대 또는 과소 포함될 위험
- 5) 군집 내의 구성원이 동질적일 경우 오차

단순 무작위 표본추출

SRS: Simple Random Sampling

모집단에서 추출한 소수의 표본이 전체 모집단의 특성을 잘 대표해야만 한다. 표본의 대표성을 담보할 수 있는 표본추출방법인 무작위 추출(random sampling)이란 모집단에 속한 개체들로부터 무작위로 일정 수의 개체를 추출하여 표본으로 구성하는 방법이다.

무작위 표본추출은 모집단을 구성하는 모든 개체들에 대하여 표본으로 뽑힐 수 있는 확률을 동일하게 부여할 수 있는 확률표본추출(probability sampling) 방법이다.

단순 무작위 표본추출(simple random sampling)은 대표적인 확률 표본추출 방법으로 모집단의 모든 구성원을 나열한 표본추출 프레임 (sampling frame) 으로부터 연구자의 인위적인 개입을 배제하면서 무작위로 표본을 추출하는 방법이다.

단순 무작위 표본추출 방법은 물리적 혼합법(physical mix)과 난수 이용법(random digits method)이 있다.

난수표를 이용한 SRS

SRS with random digits

난수표(random digits table)란 자릿수에 무관하게 동일한 확률을 갖도록 만들어진 숫자들이 나열된 표를 말한다. 예를 들면 0 ~ 9의 숫자는 1/10 확률, 00 ~ 99의 숫자는 1/100 확률, 000 ~ 999의 숫자는 1/1000 확률을 갖도록 만들어진 숫자들이 나열된 표이다.

난수표를 이용한 SRS

1단계: 표본추출 프레임의 각 구성원들에 동일한 자릿수의 숫자를 부여한다.
예를 들면 00, 01, 02, 03, 99 등.

2단계: 난수표로부터 추출된 번호에 해당하는 개체를 표본으로 선택한다. 이 때 표본추출 프레임의 리스트를 벗어난 숫자 또는 중복된 숫자는 배제한다.

난수표

Random digits

Line

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785