

## 응용사회학연구에서 확률표본추출의 방법

리 광 호

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《사회과학부문에서는 당면하여 응용사회학을 비롯하여 우리의 혁명실전에 필요한 분야들을 우리 식으로 개척하기 위한 과학연구사업을 적극적으로 벌리며 점차 그 범위를 넓혀나가도록 하여야 합니다.》

응용사회학은 사회를 과학적으로, 합리적으로 관리운영해나가는데서 제기되는 구체적인 실천적문제들을 정량적으로 연구하는 학문이다.

응용사회학연구에서는 연구하려는 모집단에서 일부 대상들을 표본으로 뽑고 그 표본에 대한 연구를 통하여 모집단을 파악하는 방법을 많이 적용한다.

모집단에서 표본대상들을 뽑아내는 방법은 그 기초원리에 따라 크게 확률표본추출과 비확률표본추출의 두가지 유형으로 가른다. 확률표본추출은 확률론의 기본원리에 근거하여 우연성의 원칙에 따라 표본을 뽑아내는 추출이며 비확률표본추출은 주로 연구자의 주관적인 희망이나 판단 혹은 편리성에 따라 대상을 뽑아내는 추출이다.

응용사회학연구에서 쓰이는 확률표본추출방법들은 우선 단순우연표본추출방법이다.

단순우연표본추출방법은  $N$ 개의 원소들을 포함하고있는 모집단속에서 직접 우연적으로  $n$ 개의 원소를 뽑아( $N > n$ 인 조건에서) 표본을 만드는 추출방법이다.

모집단의 규모  $N$ 이 크지 않은 경우 단순우연표본추출방법은 다음과 같다. 먼저 모집단의 매개 원소(즉 표본추출단위)들의 명단(즉 표본추출틀)을 작성한 다음 매개의 원소들을 부호화하고 그 부호들을 종이쪽지에 하나씩 쓴다. 그다음 모든 종이쪽지들을 통에 담고 뒤섞은 다음 그중에서 임의로  $n$ 개의 쪽지를 뽑아낸다. 이렇게 하여 뽑혀나온 번호에 의하여 대표되는 원소들로 구성된것이 바로 단순우연표본이다.

실례로 어떤 단위에 200명의 사람이 있는데 단순우연표본추출의 방법으로 20명의 사람들을 뽑아 표본을 만들려고 한다고 하자. 그러면 먼저 200명의 명단을 작성한 다음 매 사람에게 001부터 200까지 부호를 하나씩 매긴다.(표본추출틀준비) 그다음 200장의 종이쪽지에 001부터 200번까지 하나씩 번호를 적어넣는다. 번호가 적혀있는 200장의 종이쪽지를 통안에 넣고 흔들어 고루 섞이게 한 다음 임의의 20장의 쪽지를 꺼낸다. 다음에 꺼낸 20장의 종이쪽지에 씌여진 번호에 대응하는 20명의 사람들을 찾아낸다. 이 20명이 바로 단순우연표본추출방법으로 뽑은 표본으로 된다. 이 방법은 모집단의 원소수가 비교적 적을 때 리용된다.

모집단의 원소수가 많은 경우에는 란수를 리용하여 다음과 같이 표본추출을 한다. 먼저 모집단의 모든 원소들의 명단 즉 표본추출틀을 준비한 다음 순서대로 부호화한다. 그다음 모집단의 규모가 몇자리수인가에 따라 몇자리수의 란수를 만들겠다는가를 확정한다. 그리고 컴퓨터로 표본크기에 해당하는  $n$ 개의 란수를 발생시키며 표본틀에서 그에 해당하는 번호의 대상들을 찾아내어 표본으로 삼는다.

실례로 5만명의 사람들로 구성된 모집단에서 200명을 표본으로 뽑아 연구를 진행해야 한다고 하자. 우선 모집단성원들의 명단을 작성한 다음 매개 사람들에게 1부터 50 000까지 부호를 붙인다. 그다음 모집단의 규모(5만명)가 5자리수이므로 컴퓨터로 5자리수의 란수들을 발생시킨다. 이때 발생된 란수들가운데서 5만보다 큰 수자들은 버리고 남은 수자가 200개가 될 때까지 란수를 발생시킨다. 발생된 번호에 따라 모집단명단에서 그에 대응하는 200명을 찾아낸다. 이 200명의 성원들이 표본으로 된다.

단순우연표본추출방법은 조작이 간단하고 표본의 우연성을 높일수 있는 우점을 가지고있다.

응용사회학연구에서 쓰이는 확률표본추출방법은 또한 분층표본추출방법이다.

분층표본추출방법은 류형표본추출방법이라고도 하는데 먼저 일정한 징표에 따라 모집단을 호상 배제적인 몇개의 부분모집단(류형)들로 가르른 다음 매개의 부분모집단들에서 각각 부분표본들을 뽑아내고 그 부분표본들을 합쳐 모집단의 표본을 만드는 추출방법이다. 사회학에서는 일정한 징표에 따라 모집단에서 갈라져나온 부분모집단을 층이라고 하며 모집단을 층으로 가르고 표본을 만든다는 의미에서 이 방법을 분층표본추출방법이라고 한다.

실례로 나라의 모든 공장, 기업소들의 기업관리실태를 료해하기 위하여 먼저 공장, 기업소들을 인민경제부문별로 가르고 매개 부문마다에서 임의로 한개 공장을 선택하여 실태를 료해하는 방법은 분층표본추출방법을 적용한것이다. 그리고 수산협동조합들의 실태를 료해하기 위하여 먼저 수산협동조합들을 도별로 가르고 매개 도에서 4~5개의 수산협동조합들을 선택하는것도 분층표본추출방법을 리용한것이다.

다른 실례로 어느 한 도의 청년사업일군들이 자기 도내의 청년들속에서 첨단과학기술을 소유하기 위하여 얼마나 노력하는가 하는것을 료해하고 대책하러 한다고 하자. 이때 해당 도안의 모든 청년들이 모집단으로 되는데 이 모집단은 성별, 직업, 지식정도 등 여러 징표들에서 서로 다른 각이한 특성을 가진 수많은 청년들로 이루어지게 된다. 이가운데서 직업 하나만 놓고보아도 서로 다른 직업을 가진 청년들속에서 첨단과학기술을 소유하려는 노력이 똑같이 나타날수는 없을것이다. 그러므로 실태를 정확하게 료해하려면 청년들을 직업에 따라 로동자, 농민, 지식인, 대학생, 사무원 등 여러가지 부류(층)로 나누고 매개 부류마다에서 각각 표본을 추출한 다음 그것을 종합하여 모집단의 전체 표본으로 삼을수 있다.

일반적으로 분층표본추출을 할 때 층으로 갈라져나온 부분모집단들의 규모의 비례와 부분표본들의 크기비례가 같게 하는것이 좋다.

실례로 어떤 공장에 로동자가 600명 있는데 남성로동자가 500명, 녀성로동자가 100명으로서 그 비가 5:1이라고 하자. 만일 60명을 뽑아 표본을 만들려고 한다면 모집단의 성비에 맞게 남성로동자중에서 50명, 녀성로동자중에서 10명을 우연추출하면 표본의 성별구조는 모집단에서의 성별구조와 똑같게 되어 성별구조에서의 표본의 대표성이 높아지게 된다.

분층표본추출방법을 리용하면 표본오차를 줄이고 표본추출의 정확도를 높일수 있다. 모집단의 동질성이 높을수록 표본의 대표성이 높아지며 모집단의 이질성이 높으면 높을수록 대표성이 낮아진다. 분층표본추출의 방법을 리용하면 이질성이 강한 모집단을 동질성이 강한 부분모집단들로 갈라 표본추출의 효률을 높이고 보다 좋은 표본추출효과를 얻을수 있다. 분층표본추출방법을 리용하면 모집단내의 각이한 분층들의 실태도 쉽게 료해할수 있다.

응용사회학연구에서 쓰이는 확률표본추출방법은 또한 집단표본추출방법이다.

집단표본추출방법은 모집단속의 개별적인 원소가 아니라 개별적인 원소들로 이루어진 일정한 크기의 집단을 표본추출단위로 하여 표본을 만드는 추출방법이다. 다시말하여 집단표본추출방법은 단순우연표본추출방법이나 분층표본추출방법에서와 같이 개별적인 원소들을 표본추출단위로 하여 모집단이나 부분모집단에서 직접 개별적인 대상들을 표본으로 뽑는것이 아니라 모집단에 속한 소집단들을 대상으로 그가운데서 몇개의 집단을 뽑고 거기에 속한 개별적대상들을 모두 표본으로 삼는 방법이다.

실례로 농장의 개별적인 농촌집들에서 닭을 5마리씩 기르는 시험을 하기 위하여 직접 농장안의 개별적인 농촌집들을 선택하는것이 아니라 먼저 어느 하나의 작업반을 선택하고 그 작업반에 속한 모든 농촌집들을 시험농가로 뽑는것은 집단표본추출방법을 적용한것이다.

다른 실례로 한 학급에 20명씩 100개 학급이 있는 대학에서 100명의 학생들을 표본으로 뽑기 위하여 100개의 학급에서 임의의 5개의 학급을 뽑고 그 5개 학급에 속한 학생들 100명을 표본으로 삼는 방법이 집단표본추출방법이다.

집단표본추출방법을 리용하면 표본추출과정을 간단히 할수 있다. 단순우연표본추출이나 분층표본추출을 하기 위하여서는 모집단의 모든 성원들에 대한 명단 즉 표본추출틀이 있어야 한다. 그러나 실제로 그러한 명단을 얻기가 힘든 경우가 많다. 또 어떤 경우에는 표본틀을 얻을수는 있지만 진짜 써먹자면 매우 불편한 때도 있다.

실례로 10만세대의 가정이 있는 도시에서 1 000세대의 가정을 조사하려 할 때 위의 두가지 방법으로 한다면 반드시 먼저 이 10만세대의 명단을 만들어야 한다. 그런데 실제로 이러한 명단을 만들기는 힘들다. 이때 집단표본추출방법을 적용하면 표본추출을 간단하고 쉽게 할수 있다. 집단표본추출방법은 모집단성원들이 학급이나 작업반과 같이 소규모집단을 이루고있는 경우에 리용할수 있다.

응용사회학연구에서 쓰이는 확률표본추출방법은 또한 다계단표본추출방법이다.

다계단표본추출방법은 모집단의 원소들을 일정한 징표에 따라 몇개의 준위로 분류하고 매개 준위마다에서 표본을 만드는 추출방법이다.

실례로 보통교육부문에서 전국의 고급중학교들의 실험교육조건을 료해하고 필요한 대책을 세우려고 한다고 하자. 이때 먼저 전국의 고급중학교들을 도별로 분류하고 몇개의 도를 선택한다.(제1단계 표본추출) 그다음 선택된 도들에서 다시 몇개의 군을 선택한다.(제2단계 표본추출) 끝으로 선택된 군들에서 몇개의 고급중학교들을 선택한다.(제3단계 표본추출) 이런 방법으로 뽑혀진 중학교들이 바로 전체 모집단의 표본으로 되며 이러한 방식으로 표본을 만드는 방법이 바로 다계단표본추출방법이다.

다계단표본추출방법은 모집단의 범위가 특별히 크고 대상들의 준위가 특별히 많은 경우에 적용한다. 다계단표본추출방법에서는 모집단의 전체 명단이 필요없으며 매 단계의 표본추출단위수가 일반적으로 좀 작기때문에 표본을 추출하기가 비교적 쉽다. 그러나 매 단계의 표본추출때마다 오차가 생겨날수 있는데 이러한 표본추출방법의 오차는 비교적 클수 있다. 다계단표본추출에서 오차를 줄이기 위해서는 시작단계의 표본수는 상대적으로 늘이고 맨 마지막단계의 표본수는 적당히 줄여야 한다.

응용사회학연구에서 쓰이는 확률표본추출방법은 또한 체계적표본추출방법이다.

체계적표본추출방법은 등거리표본추출방법이라고도 하는데 표본추출단위들의 명단 즉 표본추출틀에서 일정한 간격마다 하나씩 체계적으로 뽑아내어 표본을 만드는 추출방법이다.

체계적표본추출을 할 때에는 우선 모집단안의 모든 개별적대상들의 명단을 작성하고 번호를 매긴다. 즉 표본추출틀을 만든다. 그다음 공식  $K(\text{표본추출간격})=N(\text{모집단규모})/n(\text{표본규모})$ 을 리용하여 표본추출간격을 계산해낸다. 그리고 표본추출틀에서 제일 앞에 있는 K개의 대상들가운데서 단순우연표본추출의 방법으로 하나의 대상을 뽑는다. 끝으로 그 대상의 번호로부터 매 K번째 번호의 대상들을 뽑아 표본을 만든다.

실례로 대학생이 3 000명인 대학에서 50명의 대학생을 뽑아 표본을 만든다고 하자.

우선 3 000명의 대학생들의 명단을 작성하고 순서대로 번호를 매긴다. 그다음 위의 공식에 따라 표본추출간격을 계산한다. 그러면  $K=3\,000/50=60$  즉 매 60명당 1명씩 뽑아내야 한다. 그러므로 1부터 60까지의 번호가운데서 단순우연표본추출의 방법으로 하나의 수자를 뽑아낸다. 가령 32가 나왔다고 하면 이 32를 첫 번호로 하여 매 60명 간격으로 한명씩 뽑아낸다. 그러면 32, 92, 152, ..., 2 972 총 50개의 번호를 뽑아내게 되는데 그 번호에 해당되는 50명의 학생들이 곧 표본으로 되게 된다.

체계적표본추출방법은 첫 표본추출간격에서 하나의 대상만 우연표본추출하기때문에 확률표본추출방법들가운데서 가장 쉬운 표본추출방법이라고 말할수 있다.

실천에서는 우와 같은 5가지 표본추출방법들을 단독으로 사용하는것이 아니라 여러가지 표본추출방법들을 적절히 배합하여 표본을 만든다.

실례로 전국의 고급중학교의 교육조건실태를 다계단표본추출의 방법으로 조사한다고 할 때 첫 단계에서는 단순우연표본추출의 방법으로 몇개의 도를 선정하고 두번째 단계에서는 분층표본추출의 방법으로 산골군과 벌방군, 도시를 선택하며 다음 선택된 군들에서는 체계적표본추출방법으로 몇개 고급중학교를 선택하며 선택된 학교들에서 집단표본추출방법으로 몇개 학급을 표본으로 선택할수 있다.

우리는 확률표본추출방법을 잘 알고 그것을 사회현상연구에 적극 활용하여 우리의 혁명실천에서 해결을 기다리는 여러가지 사회적문제들을 과학리론적으로 깊이있게 해명함으로써 사회주의강국건설위업수행에 적극 이바지하여야 할것이다.