

11강. 집락추출법(2)

◆ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
규모측도(measure of size : MOS)	확률비례추출법에서 표본추출을 위해서 채택된 측도를 말함. 일반적으로 조사항목과 상관관계가 높은 것을 채택하면 추정의 효율이 높아짐
확률비례추출법 (Probability Proportional to Size Sampling : PPS Sampling)	모집단을 구성하는 집락들의 규모가 심하게 차이 날 경우에 동일한 확률로 뽑지 않고 각 집락의 규모측도에 비례하여 뽑는 방법

■ 실습하기

- 교재 195쪽 예제 6-4
 - * 집락추출법(집락의 크기가 서로 다른 경우) 사례
 - * 모평균 추정, 표준오차, 신뢰구간 작성

■ 연습문제

1. 집락추출법에서 추출단위인 집락의 규모가 심하게 차이 나고, 조사변수의 집락 합계와 집락의 크기가 강한 상관관계를 나타내고 있다. 이러한 경우에 추정의 정확도를 높일 수 있는 표본추출방법은 다음 중 어느 것인가?
 - ① 각 집락의 크기에 반비례하여 표본집락을 추출한다.
 - ② 각 집락의 크기에 비례하여 표본집락을 추출한다.
 - ③ 단순임의추출법에 따라 표본집락을 추출한다.
 - ④ 계통추출법에 의해서 표본집락을 추출한다.

정답 : ②

해설 : 확률비례추출법은 조사변수의 합계와 집락 내의 조사단위의 수와 양의 상관관계가 뚜렷할 때 효율적이다.

2. 이 지역 내에 있는 색맹인 초등학교 1학년 학생의 총수를 추정하고, 95% 신뢰구간을 구하라.

※(2~3) 어느 지역의 교육당국은 초등학교 1학년 학생들을 대상으로 색맹인 학생의 총수와 그 비율을 추정하려고 한다. 이 지역에 있는 전체 초등학교 수는 30개이고, 이들 중에서 5개 초등학교를 추출하여 다음 결과를 얻었다.

학교번호	1	2	3	4	5
초등학교1학년 학생수	130	240	110	180	100
색맹인 학생 수	2	3	2	6	0

정답 : 이 지역 내의 총 초등학생 수(N)를 알 수 없다. 따라서 이 지역에서 색맹인 초등학교 1학년 학생의 총수는 다음 식을 통해서 구할 수 있다.

$$\hat{\tau} = A\bar{y}_t = A \times \frac{\sum_{i=1}^a y_i}{a} = 30 \times \frac{13}{5} = 78$$

$$\hat{V}(\hat{\tau}) = A^2 \hat{V}(\bar{y}_t) = A^2 \left(\frac{A-a}{A} \right) \frac{1}{a} \frac{\sum_{i=1}^a (y_i - \bar{y}_t)^2}{a-1} = 30^2 \left(\frac{30-5}{30} \right) \frac{1}{5} \frac{19.2}{4} = 720$$

$$\begin{aligned} \text{- 95\% 신뢰구간 : } \hat{\tau} \pm \sqrt{\hat{V}(\hat{\tau})} &\Leftrightarrow 78 \pm 2\sqrt{720} \\ &\Leftrightarrow 78 \pm 53.67 \end{aligned}$$

3. 이 지역 내에 있는 초등학교 1학년 학생 중 색맹인 학생의 비율을 추정하고, 이에 대한 95% 신뢰구간을 구하라.

※(2~3) 어느 지역의 교육당국은 초등학교 1학년 학생들을 대상으로 색맹인 학생의 총수와 그 비율을 추정하려고 한다. 이 지역에 있는 전체 초등학교 수는 30개이고, 이들 중에서 5개 초등학교를 추출하여 다음 결과를 얻었다.

학교번호	1	2	3	4	5
초등학교1학년 학생수	130	240	110	180	100
색맹인 학생 수	2	3	2	6	0

정답 : $\hat{p} = \sum_{i=1}^a y_i / \sum_{i=1}^a B_i = 13/750 = 0.0173$

$$\hat{V}(\hat{p}) = \left(\frac{A-a}{A} \right) \frac{1}{\bar{b}^2} \frac{1}{a} \frac{\sum_{i=1}^a (y_i - \hat{p} B_i)^2}{a-1}$$

$$= \left[\frac{50-5}{50} \right] \frac{1}{150^2} \frac{1}{5} \left(\frac{12.35}{4} \right)$$

$$= 0.000124$$

따라서 이 지역 내의 색맹인 학생의 비율 대한 95% 신뢰구간은 다음과 같다.

$$\hat{p} \pm 2\sqrt{\hat{V}(\hat{p})} \leftrightarrow 0.0173 \pm 2\sqrt{0.000124}$$

4. 다음은 집락추출법에 대한 설명이다. 옳은 것끼리 짝지어진 것은?

- ㉠ 집락 내 단위들은 동질적이고, 집락 간의 변동이 큰 경우에 효과적이다.
 ㉡ 모집단에 대한 완전한 추출률이 마련되지 않은 경우에도 사용할 수 있다.
 ㉢ 대개 집락추출법을 적용하면 설계효과(DEFF)는 1보다 크게 나타난다.

- ① ㉠, ㉡
 ② ㉠, ㉢
 ③ ㉡, ㉢
 ④ ㉠, ㉡, ㉢

정답 : ③

해설 : 집락추출법은 집락 내에서는 이질적이고, 집락 간에는 동질적인 경우에 추정
 효율이 높다.

■ 정리하기

1. 모평균 추정 : 집락의 크기가 다른 경우

$$\bar{y}_{cl} = \frac{\sum_{i=1}^a y_i}{\sum_{i=1}^a B_i}$$

$$\hat{V}(\bar{y}_{cl}) = \left(\frac{A-a}{A} \right) \frac{1}{\bar{B}^2} \frac{1}{a} \frac{\sum_{i=1}^a (y_i - \bar{y}_{cl} B_i)^2}{a-1}$$

2. 확률비례추출법은 집락이 추출될 확률이 집락의 크기에 비례하는 추출법이며, 조사변수의 집락의 합계와 집락 내의

조사단위수간에 높은 상관관계를 가질 때 효율적이다.

3. 확률비례추출법에서 모평균과 모총계 추정

- A 개 집락으로 구성된 모집단에서 a 개 집락 추출

- 사전에 알려진 집락의 크기 B_i 에 비례하여 복원확률비례추출하는 경우

- 모총계 추정

$$\hat{\tau}_{PPS} = \frac{N}{a} \sum_{i=1}^a \frac{y_i}{B_i} = \frac{N}{a} \sum_{i=1}^a \bar{y}_i \text{ 여기서, } y_i = \sum_{j=1}^{B_i} y_{ij}, \quad \bar{y}_i = \frac{y_i}{B_i}$$

- 모평균 추정

$$\hat{\mu}_{PPS} = \frac{\hat{\tau}_{PPS}}{N} = \frac{1}{a} \sum_{i=1}^a \bar{y}_i$$

$$\widehat{V}(\hat{\mu}_{PPS}) = \frac{1}{a(a-1)} \sum_{i=1}^a (\bar{y}_i - \hat{\mu}_{PPS})^2$$

■ 참고문헌

- 이계오, 박진우, 이기재, 표본조사론, 한국방송통대학교출판부, 2013. 제1장
- 통계청 홈페이지 : <http://www.nso.go.kr>