

8강. 층화임의추출법(3)

◆ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
네이만배분법(Neyman allocation)	각 층 내의 모집단 추출단위들의 수와 층별 변동을 동시에 고려하여 최적의 표본의 크기를 배분하는 방법
최적배분법(optimum allocation)	각 층 내의 모집단 추출단위들의 수와 층별 변동 및 층별 조사비용을 동시에 고려하여 최적의 표본의 크기를 배분하는 방법
사후층화(post-stratification)	단순임의추출법에 따라 얻어진 표본의 구성이 이미 알고 있는 모집단 특성 비율을 제대로 반영하지 못할 경우 표본 데이터에 대하여 사후적으로 층을 구분하여 추정하는 방법

■ 실습하기

- 교재 128쪽 예제 4-7, 예제 4-8, 예제 4-9
 - * 비례배분법, 네이만배분법, 최적배분법 적용
- 교재 134쪽 예제 4-10
 - * 사후층화 사례

■ 연습문제

1. 다음은 층화임의추출에서 각 층에 표본을 배분할 때 고려할 요인들이다. 알맞지 않은 것은?
 - ① 각 층의 크기(N_h)
 - ② 각 층에서의 변동(S_h^2)
 - ③ 각 층에서의 단위당 조사비용(C_h)
 - ④ 신뢰수준

정답 : ④

해설 : 신뢰수준은 표본의 크기 결정에 영향을 준다.

2. 다음은 층화임의추출법에서 각 층에 표본을 배분하는 방법에 대한 설명이다. 옳지 않은 것은?

- ① 비례배분법을 적용하면 각 추출단위에 대한 추출률이 모두 같게 된다.
- ② 비례배분법은 층별 변동의 크기에 비례해서 표본을 배분하는 방법이다.

$$n_h = n \frac{N_h S_h}{\sum_{k=1}^H N_k S_k}$$

- ③ 네이만 배분법은 각 층에 $\sum_{k=1}^H N_k S_k$ 과 같이 배분한다.
- ④ 최적배분법은 주어진 비용 하에서 추정량의 분산을 최소화시키는 방법이다.

정답 : ②

해설 : 비례배분법은 층별 모집단 크기에 비례해서 표본을 배분하는 방법이다.

3. 가구 수가 각각 인 세 도시를 대상으로 표본조사를 실시하여 월 평균 가구소득을 조사하고자 한다. 전체 표본크기를 n=100가구로 한다.

각 도시별 표준편차가 각각 인 것이 알려져 있다고 한다. 네이만배분법에 의해 도시별 표본의 크기를 배분하여라.

정답 및 해설 : 각 층별 모집단의 크기와 층별 변동을 고려하는 네이만배분법의 공식은 다음과 같다.

$$n_h = n \left(\frac{N_h S_h}{\sum_{k=1}^H N_k S_k} \right), \quad h = 1, 2, \dots, H$$

세 도시에 대한 네이만배분법에 의한 표본배분 결과는 각각 45가구, 18가구, 37가구이다.

4. 이 문제에 대해서 층화임의추출법을 적용하는 대신에 같은 표본크기인 100가구를 단순임의추출법으로 추출하여 표본평균으로 월 평균 생계비를 추정하였다. 다음의 설명 중 옳은 것은?

- ① 층화임의추출법을 적용하는 것에 비해서 추정의 효율이 높아진다.
- ② 층화임의추출법을 적용하는 것에 비해서 추정량에 편향이 커진다.
- ③ 층화임의추출법을 적용하는 것에 비해서 추정의 정확도가 떨어질 것이다.
- ④ 층화임의추출법을 적용하는 것에 비해서 추정량에 편향이 작아진다.

정답 : ③

해설 : 단순임의추출법이나 층화임의추출법은 확률추출법으로 이들 방법을 적용할 때 추정에 편향은 발생하지 않는다. 일반적으로 층화임의추출법은 단순임의추출

법에 비해서 추정의 효율이 높아서 층화임의추출법을 적용하는 것이 단순임의추출법에 비해서 추정의 정확도가 높다.

5. 다음과 같은 표본조사 결과를 기초로 이 도시의 전체 흡연율을 사후층화 방법을 적용하여 추정하면 추정값은 얼마인가?

인구 10만인 도시가 있는데, 이 도시에 거주하는 남자와 여자의 비율이 각각 50%이다. 이 도시 전체 시민의 흡연율을 추정하기 위해 단순임의추출법으로 1,000명을 추출했는데 이 표본에 남자가 800명, 여자가 200명 뽑혔으며 표본에 뽑힌 남자와 여자의 흡연율은 각각 50%와 20%였다고 하자.

- ① 44%
- ② 50%
- ③ 35%
- ④ 20%

정답 : ③

해설 :

$$\hat{p}_{post} = \sum_{h=1}^H W_h \cdot \hat{p}_h = 0.5 \times 0.5 + 0.5 \times 0.2 = 0.35$$

■ 정리하기

1. 층화추출법에서 전체 표본의 크기를 각 층별로 배분하는 방법은 중요한 관심사이다.

각 층별 모집단의 크기와 층별 변동을 고려하는 **네이만배분법**의 공식은 다음과 같다.

$$n_h = n \left(\frac{N_h S_h}{\sum_{k=1}^H N_k S_k} \right), \quad h = 1, 2, \dots, H$$

2. 층화추출법에서 각 층별 모집단의 크기와 층별 변동 및 층별 조사비용을 고려하는 **최적배분법**의 공식은 다음과 같다.

$$n_h = n \left(\frac{N_h S_h / \sqrt{c_h}}{\sum_{k=1}^H N_k S_k / \sqrt{c_k}} \right), \quad h = 1, 2, \dots, H$$

3. 총화는 표본설계 단계에서 모집단의 모든 단위들에 대해 사전에 정보를 알고 있는 총화변수를 기준으로 하여 이루어지는데 반해

사후총화는 표본추출이 이루어지고 난 이후 조사된 데이터에 근거하여 표본의 단위들을 총화하는 것이다.

사후총화를 이용한 모평균의 추정량 및 그 분산추정량은 다음과 같다.

$$\bar{y}_{post} = \sum_{h=1}^H W_h \cdot \bar{y}_h$$

$$\hat{V}(\bar{y}_{post}) = \frac{N-n}{Nn} \cdot \sum_{h=1}^H W_h s_{ph}^2 + \frac{1}{n^2} \sum_{h=1}^H (1 - W_h) s_{ph}^2$$

4. 효과적으로 총화를 할 경우 총화임의추출법은 단순임의추출법에 비해 매우 효율적이다.

하지만 총화가 적절하게 이루어지지 못한다면 오히려 단순임의추출법보다 효율이 떨어질 수도 있다.

총화변수가 조사하고자 하는 주변수와 상관이 높을수록 효과적인 총화가 된다.

■ 참고문헌

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