

## 6강. 층화임의추출법(1)

◆ 담당교수 : 이기재 교수

### ■ 주요용어

용어	해설
층(stratum)	모집단의 조사단위들을 특성에 따라 서로 겹치지 않는 몇 개의 부차모집단(subpopulation)으로 나누었을 때 각각의 부차모집단을 일컬음
층화	모집단을 어떤 특성에 따라 몇 개의 부차모집단으로 나누는 일
층화변수	모집단을 층화할 때 층화의 기준으로 삼는 변수. 관심변수와 상관성이 높은 변수를 선택하는 것이 바람직함
층화임의추출법	모집단을 먼저 서로 겹치지 않는 여러 개의 층으로 분할한 후, 각 층별로 단순임의추출법을 적용시켜 표본을 얻는 방법
누적도수제곱근법	층의 최적경계점을 정하기 위해 가장 널리 쓰이는 근사적인 방법으로 층화변수 값에 따른 누적도수분포표를 작성하여 누적도수의 총계를 층의 수로 나눈 값의 배수를 층의 경계점으로 정하는 방법

### ■ 실습하기

- 교재 117쪽 예제 4-3
  - \* 층화임의추출법에서 모평균 추정
- 교재 143쪽 연습문제 8번
  - \* 층화임의추출법에서 모평균 추정

### ■ 연습문제

1. 모집단을 서로 겹쳐지지 않는 몇 개의 부분군으로 나누었을 때 각각의 부분군들을 일컬어 ( )이라고 한다.  
이때 모집단을 서로 겹쳐지지 않게 몇 개의 부분군으로 나누는 일을 ( )라고 한다.

정답 및 해설 : 층(stratum), 층화

2. 다음은 층화임의추출법에 대한 설명이다. 옳지 않은 것은?

- ① 모집단 전체에 대한 추정뿐만 아니라 각 층별 추정도 가능하다.
- ② 모집단 내의 모든 추출단위들은 같은 추출확률을 갖는다.
- ③ 같은 층에 속한 기본단위들은 동질적이고, 층간에는 이질적일 때 효과적이다.
- ④ 조사관리 측면에서 단순임의추출법에 비해서 편리하다.

정답 : ②

해설 : 층화임의추출법에서 비례배분법을 적용하면 모집단 내의 모든 추출단위가 같은 추출확률을 갖게 되지만 일반적으로 층에 따라 추출확률이 달라지게 된다.

3. 점포규모가 소(小 : 층 1)에 대한 하루 평균 판매액에 대한 95% 신뢰구간을 구하면?

$$\text{단, } \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \hat{V}(\bar{y}_h) = \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$$

※(3~4) N=500의 슈퍼마켓을 점포면적 규모로 층화하여 n=50의 표본을 각 층에 비례배분시켜 하루당 판매액을 조사한 결과가 다음의 표와 같다.

층	$N_h$	$n_h$	$\bar{y}_h$	$s_h^2$	$\frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$
1(소)	300	30	27	114.2	3.4
2(중)	150	15	59	555.9	33.4
3(대)	50	5	124	2444.7	440.1
계	500	50			

- ①  $27 \pm 3.7$
- ②  $27 \pm 5.7$
- ③  $59 \pm 3.7$
- ④  $59 \pm 5.7$

정답 : ①

해설 :

$$\hat{V}(\bar{y}_1) = \frac{N_1 - n_1}{N_1} \frac{s_1^2}{n_1} = 3.4, \quad \bar{y}_1 \pm z_{\alpha/2} \sqrt{\hat{V}(\bar{y}_1)} \Leftrightarrow 27 \pm 2 \sqrt{3.4} \Leftrightarrow 27 \pm 3.7$$

4. 전체 슈퍼마켓의 하루 평균 판매액에 대한 추정치는 얼마인가?

$$(\text{단, } \bar{y}_{st} = \frac{1}{N} \sum_{h=1}^H N_h \bar{y}_h)$$

※(3~4) N=500의 슈퍼마켓을 점포면적 규모로 층화하여 n=50의 표본을 각 층에 비례배분시켜 하루당 판매액을 조사한 결과가 다음의 표와 같다.

층	$N_h$	$n_h$	$\bar{y}_h$	$s_h^2$	$\frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$
1(소)	300	30	27	114.2	3.4
2(중)	150	15	59	555.9	33.4
3(대)	50	5	124	2444.7	440.1
계	500	50			

- ① 112.3
- ② 99.3
- ③ 69.3
- ④ 46.3

정답 : ④

해설 :

$$\begin{aligned} \bar{y}_{st} &= \frac{1}{N} \sum_{h=1}^L N_h \bar{y}_h \\ &= \frac{1}{500} (300 \times 27 + 150 \times 59 + 50 \times 124) = 46.3 \end{aligned}$$

5. 다음은 층화임의추출법에 대한 설명이다. 옳은 것을 모두 고른 것은?

가. 각 층 내부의 조사단위가 동질적이고, 층들 간에는 이질적일 때 효과적이다.  
 나. 전체 모집단에 대한 추정뿐만 아니라 각 층별 추정도 가능하다.  
 다. 최적배분법을 적용하면 모집단 내의 모든 추출단위들은 같은 추출확률을 갖는다.  
 라. 단순임의추출법에 비해서 조사관리가 편리하고, 조사비용도 절감할 수 있다.

- ① 가, 나, 라
- ② 가, 다, 라
- ③ 나, 다
- ④ 나, 라

정답 : ①

해설 : 임의추출법에서 비례배분법을 적용하면 모집단 내 모든 추출단위의 추출확률

이 같게 된다.

## ■ 정리하기

- 모집단을 서로 겹쳐지지 않는 몇 개의 부분군으로 나누었을 때 각각의 부분군들을 일컬어 층(stratum)이라고 한다. 층화임의추출법이란 모집단을 몇 개의 층으로 나눈 후 각 층에 단순임의추출법을 적용시켜 표본을 추출하는 방법이다.
- 모집단을 층으로 나눌 때 기준으로 삼는 층화변수가 조사하고자 하는 관심변수와 상관이 높을수록 층화는 효과적이다. 같은 층에 속한 기본단위들은 서로 동질적인 값을 갖고, 층들 간에는 서로 이질적일수록 층화의 효율은 높아지게 되는데 이렇게 효율을 높일 수 있으면서도 편리한 층화변수를 마련하는 것이 바람직하다.
- 층화임의추출법의 장점은, 전체 모집단의 추정뿐만 아니라 각 층에 대한 추정이 가능하여 층별 통계 또는 지역통계를 생산할 수 있다는 점과, 단순임의추출법에 비해 조사관리가 편리하며 조사비용도 절감할 수 있다는 점이다. 층화임의추출법의 장점을 살리기 위해서는 적절한 층화변수를 정하는 것이 매우 중요하다.
- 층화임의추출법에서 모평균에 대한 추정량과 분산은 다음과 같다.

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^H N_h \bar{y}_h$$

---

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$$

## ■ 참고문헌

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