

7강. 층화임의추출법(2)

◆ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
표본배분(sample allocation)	층화임의추출법에서 표본의 크기가 정해졌을 때 추정의 효율을 최대화시킬 수 있도록 각 층별로 표본수를 배분하는 전략
비례배분법 (proportional allocation)	각 층 내의 모집단 추출단위들의 수에 비례하여 표본의 크기를 배분하는 방법. 일반적으로 사회조사에서 널리 활용됨

■ 실습하기

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

■ 연습문제

1. A시(층 1)의 특정 TV 프로그램 시청률에 95% 신뢰구간은?

$$\text{(단, } \hat{V}(\hat{p}_t) = \frac{N_t - n_t}{N_t} \frac{\hat{p}_t \hat{q}_t}{n_t - 1})$$

※[1-3] 어느 광고회사에서는 A시, B군으로 이루어진 어느 지역을 대상으로 특정 TV 프로그램에 대한 시청률을 조사하려고 한다. 층화임의추출법을 적용하여 다음과 같은 결과를 얻었다. 물음에 답시오.

층번호	N_h	n_h	\bar{y}_h	s_h^2
1	600	30	12	4.84
2	400	20	26	10.24

- ① 0.8 ± 0.063
- ② 0.4 ± 0.063
- ③ 0.8 ± 0.080

④ 0.4 ± 0.080

정답 : ③

해설 :

$$\hat{V}(\hat{p}_1) = \frac{N_1 - n_1}{N_1} \frac{\hat{p}_1 \hat{q}_1}{n_1 - 1} = \frac{20000 - 100}{20000} \times \frac{0.8 \times 0.2}{99} = 0.0016$$

2. 이 지역 전체 가구의 특정 TV 프로그램 시청률은?

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

※[1-3] 어느 광고회사에서는 A시, B군으로 이루어진 어느 지역을 대상으로 특정 TV 프로그램에 대한 시청률을 조사하려고 한다. 층화임의추출법을 적용하여 다음과 같은 결과를 얻었다. 물음에 답하시오.

층번호	N_h	n_h	\bar{y}_h	s_h^2
1	600	30	12	4.84
2	400	20	26	10.24

- ① 0.429
- ② 0.529
- ③ 0.629
- ④ 0.729

정답 : ③

해설 :

$$\hat{p}_{st} = \frac{1}{N} \sum_{h=1}^H N_h \hat{p}_h = \frac{1}{35000} (20000 \times 0.8 + 15000 \times 0.4) = 0.629$$

3. 이 지역 전체 가구의 특정 TV 프로그램 시청률 추정량에 대한 95% 신뢰수준에서 오차의 한계는?

※[1-3] 어느 광고회사에서는 A시, B군으로 이루어진 어느 지역을 대상으로 특정 TV 프로그램에 대한 시청률을 조사하려고 한다. 층화임의추출법을 적용하여 다음과 같은 결과를 얻었다. 물음에 답하시오.

층번호	N_h	n_h	\bar{y}_h	s_h^2
1	600	30	12	4.84
2	400	20	26	10.24

(단, $\hat{V}(\hat{p}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 \frac{N_h - n_h}{N_h} \frac{\hat{p}_h \hat{q}_h}{n_h - 1}$, 오차의 한계 = $2 \times \sqrt{\hat{V}(\hat{p}_{st})}$)

- ① 0.061
- ② 0.611
- ③ 0.031
- ④ 0.311

정답 : ①

해설 :

$$\begin{aligned} \hat{V}(\hat{p}_{st}) &= \frac{1}{N^2} \sum_{h=1}^H N_h^2 \frac{N_h - n_h}{N_h} \frac{\hat{p}_h \hat{q}_h}{n_h - 1} \\ &= \frac{1}{35000^2} \left[20000^2 \frac{20000 - 100}{20000} \frac{0.8 \times 0.2}{99} + 15000^2 \frac{15000 - 100}{15000} \frac{0.4 \times 0.6}{99} \right] \end{aligned}$$

4. 운송회사들의 보험료 총 지급액을 추정하면?

※[4-5] 1,000개의 운송회사를 대상으로 한 달 동안의 보험료 총 지급액을 조사하고자 한다. 1,000개의 회사를 규모에 따라서 2개의 층으로 나눈 후 50개 회사를 층화임의추출하여 조사한 결과를 요약하였다.

층번호	N_h	n_h	\bar{y}_h	s_h^2
1	600	30	12	4.84
2	400	20	26	10.24

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

- ① 17,600
- ② 1,760
- ③ 880

④ 7,000

정답 : ①

해설 :

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^H N_h \bar{y}_h = \frac{1}{1000} (600 \times 12 + 400 \times 26) = 17.6$$

$$\hat{\tau}_{st} = N \times \bar{y}_{st} = 1000 \times 17.6 = 17,600$$

5. 보험금 총 지급액에 대한 95% 신뢰수준에서 오차의 한계 $(2 \times \sqrt{\hat{V}(\hat{\tau}_{st})})$ 는?

※[4-5] 1,000개의 운송회사를 대상으로 한 달 동안의 보험료 총 지급액을 조사하고자 한다. 1,000개의 회사를 규모에 따라서 2개의 층으로 나눈 후 50개 회사를 층화임의추출하여 조사한 결과를 요약하였다.

층번호	N_h	n_h	\bar{y}_h	s_h^2
1	600	30	12	4.84
2	400	20	26	10.24

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

- ① 329.6
- ② 529.4
- ③ 729.4
- ④ 1630.9

정답 : ③

해설 :

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

$$= 600^2 \frac{600 - 30}{600} \frac{4.84}{30} + 400^2 \frac{400 - 20}{400} \frac{10.24}{20} = 133000$$

■ 정리하기

1. 층화임의추출법에서 **모총계에 대한 추정량과 분산**

$$\hat{\tau}_{st} = \sum_{h=1}^H N_h \bar{y}_h = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

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

2. 층화임의추출법에서 **모비율에 대한 추정량과 분산**

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

$$\hat{V}(\hat{p}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 \hat{V}(\hat{p}_h) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 \frac{N_h - n_h}{N_h} \frac{\hat{p}_h \hat{q}_h}{n_h - 1}$$

3. 층화임의추출법에서 **표본크기 산출**

$$n = \frac{\sum_{h=1}^H N_h S_h^2 / w_h}{N^2 D + \sum_{h=1}^H N_h S_h^2}$$

$$\text{여기서 } D = \begin{cases} \left(\frac{B}{z_{\alpha/2}}\right)^2 & : \text{모평균 } \mu \text{를 추정하고자 할 때} \\ \left(\frac{B}{N z_{\alpha/2}}\right)^2 & : \text{모총계 } \tau \text{를 추정하고자 할 때} \end{cases} \quad \text{이고, } B \text{는 오차의 한계이다.}$$

4. 층화임의추출법에서 전체 표본의 크기를 각 층별로 배분하는 방법은 중요한 관심사이다.

각 층별 모집단의 크기만을 고려하는 **비례배분법**의 공식은 다음과 같다.

$$n_h = n \times \frac{N_h}{N}, \quad h = 1, 2, \dots, H$$

■ **참고문헌**

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