7강 층화임의추출법(2)

정보통계학과 이기재교수

학/습/목/차

1. 모수추정 – 모총계, 모비율
2. 표본크기의 결정
3. 표본배분법 – 비례배분법
4. 엑셀을 활용한 실습

모총계의 추정

1. h 번째 층의 추정

- $\hat{\tau}_h = N_h \overline{y_h} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$

2. 모총계 τ의 추정량

- $\hat{\tau}_{st} = \sum_{h=1}^{H} \hat{\tau}_h = \sum_{h=1}^{H} N_h \overline{y_h}$ 또는 $\hat{\tau} = N \overline{y_{st}}$
- $\hat{V}(\hat{\tau}_{st}) = \sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$
- $V(\hat{\tau}_{st}) = \sum_{h=1}^{H} V(\hat{\tau}_h) = \sum_{h=1}^{H} N_h^2 \, V(\overline{y_h}) = \sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}$

3. 신뢰구간 추정

- $\hat{\tau}_{st} \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{\tau}_{st})}$

모총계의 추정

✤ 예제 4-4

▶ 모총계 추정 사례

층	N_h	n_h	$\overline{y_h}$	s_h^2
1	20	5	1.6	3.3
2	9	3	2.8	4.0
3	12	4	0.6	2.2

- $\widehat{\tau_{st}} = [20(1.6) + 9(2.8) + 12(0.6)] = 64.4$

$$\widehat{V}(\widehat{\tau_{st}}) = \left[20(20-5)\frac{3.3}{5} + 9(9-3)\frac{4.0}{3} + 12(12-4)\frac{2.2}{4}\right] = 322.8$$

$$\widehat{\tau_{st}} \pm 2\sqrt{\widehat{V}(\widehat{\tau_{st}})} \leftrightarrow 64.4 \pm 2\sqrt{322.8} \leftrightarrow (28.5,\ 100.3)$$

모비율에 대한 추정

1. 모비율 p 에 대한 수학적 표현

- $p = \frac{1}{N}\sum_{h=1}^{H} N_h p_h$, 여기서 p_h는 h 번째 층의 모비율을 뜻함

2. 모비율 p 의 추정량과 분산의 추정량

- $\widehat{p_{st}} = \frac{1}{N}\sum_{h=1}^{H} N_h \widehat{p_h}$, $\widehat{p_h} = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$
- $\hat{V}(\widehat{p_{st}}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h} \frac{\widehat{p_h}\widehat{q_h}}{n_h - 1}$

모비율에 대한 추정

예제 4-5

▶ 모비율 추정 사례

층	모집단 크기 N_h	표본 수 n_h	지지자수 $\sum_{i=1}^{n_h} y_{hi}$
남자	336만	500	200
여자	364만	500	400

1. 남자와 여자의 지지율과 분산 추정

- $\hat{p}_{\text{남자}} = \dfrac{200}{500} = 0.4$, $\hat{p}_{\text{여자}} = \dfrac{400}{500} = 0.8$
- $\hat{V}(\hat{p}_{\text{남자}}) = \dfrac{3,360,000 - 500}{3,360,000} \bullet \dfrac{0.4 \times 0.6}{500 - 1} = 4.8 \times 10^{-4}$
- $\hat{V}(\hat{p}_{\text{여자}}) = \dfrac{3,640,000 - 500}{3,640,000} \bullet \dfrac{0.8 \times 0.2}{500 - 1} = 3.2 \times 10^{-4}$

모비율에 대한 추정

예제 4-5

▶ 모비율 추정 사례

층	모집단 크기 N_h	표본 수 n_h	지지자수 $\sum_{i=1}^{n_h} y_{hi}$
남자	336만	500	200
여자	364만	500	400

2. 대통령에 대한 지지율과 지지율의 95% 신뢰구간을 추정

- $\widehat{p_{st}} = 0.48 \times \hat{p}_{\text{남자}} + 0.52 \times \hat{p}_{\text{여자}} = 0.48 \times 0.4 + 0.52 \times 0.8 = 0.61$
- $\hat{V}(\widehat{p_{st}}) = 0.48^2 \times \hat{V}(\hat{p}_{\text{남자}}) + 0.52^2 \times \hat{V}(\hat{p}_{\text{여자}})$
  $= 0.48^2 \times 4.8 \times 10^{-4} + 0.52^2 \times 3.2 \times 10^{-4} = 1.97 \times 10^{-4}$
- $\widehat{p_{st}} \pm z_{\alpha/2} \sqrt{\hat{V}(\widehat{p_{st}})} \leftrightarrow 0.61 \pm 2\sqrt{1.97 \times 10^{-4}}$
  $\leftrightarrow 0.61 \pm 0.03$

학/습/목/차

1. 모수추정 – 모총계, 모비율
2. 표본크기의 결정
3. 표본배분법 – 비례배분법
4. 엑셀을 활용한 실습

표본크기의 결정

- 전체 표본의 크기
  - $n = \sum_{h=1}^{H} n_h$
- 오차의 한계
  - $B = z_{\alpha/2}\sqrt{V(\overline{y_{st}})} = z_{\alpha/2}\sqrt{\frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h}}$
- $n_h = n \times w_h \ (h = 1, 2, \ \ldots \ , H)$ 라고 가정
  - $n = \frac{\sum_{h=1}^{H} N_h^2 S_h^2 / w_h}{N^2 D + \sum_{h=1}^{H} N_h S_h^2}, \ D = \begin{cases} \left(\frac{B}{z_{\alpha/2}}\right)^2 & ;\ \mu\text{추정} \\ \left(\frac{B}{N z_{\alpha/2}}\right)^2 & ;\ \tau\text{추정} \end{cases}$

표본크기의 결정

예제 4-6

▶ 표본크기 결정 사례 : 대학생 주당 TV 시청시간 조사

- $N_1 = 155, N_2 = 62, N_3 = 93,\ \ S_1^2 \approx 25, S_2^2 \approx 225, S_3^2 \approx 100$
- $B = 2\text{시간} \Rightarrow D = \left(\dfrac{B}{2}\right)^2 = \left(\dfrac{2}{2}\right)^2 = 1\ ,\ w_h = \dfrac{1}{3}$
- $$n = \frac{\sum_{h=1}^{H} N_h^2 S_h^2 / w_h}{N^2 D + \sum_{h=1}^{H} N_h S_h^2}$$
$$= \frac{(155^2 \times 25 + 62^2 \times 225 + 93^2 \times 100)/\dfrac{1}{3}}{310^2 \times 1 + (155 \times 25 + 62 \times 225 + 93 \times 100)}$$
$$\fallingdotseq 57$$
- 층별 표본의 수 : $n_1 = n_2 = n_3 = 19$

학/습/목/차

1. 모수추정 – 모총계, 모비율
2. 표본크기의 결정
3. 표본배분법 – 비례배분법
4. 엑셀을 활용한 실습

표본의 배분

- 표본배분에 영향을 미치는 요인
  - 각 층 내의 추출단위들의 수 : (N_h)
  - 각 층 내에서 변동의 정도 : (S_h^2)
  - 각 층에서 추출단위를 조사하는데 드는 비용 : (c_h)
- 표본배분의 일반적인 원칙
  - 층 내의 추출단위 수가 많을수록 표본을 많이 배분
  - 층 내의 단위들이 이질적이어서 S_h^2 이 클수록 표본을 많이 배분
  - 조사비용이 많이 드는 층에 대해서는 가능하면 표본을 적게 배분
- 표본배분법의 종류
  - 비례배분법, 네이만배분법, 최적배분법

(1) 비례배분법

- 각 층 내의 추출단위 수(N_h)에 비례하여 표본크기를 배분하는 방법
- 층 내의 변동과 조사비용은 고려하지 않고 층의 크기만을 고려한 방법
- 층별 변동에 차이가 없고 층별로 조사비용이 비슷한 경우에 알맞음
- 일반적으로 여론조사, 의식조사 등에 많이 활용됨
- 배분공식 : $n_h = n \times \dfrac{N_h}{\sum_{h=1}^{H} N_h}$, $h = 1, 2, \ldots, H$
- 비례배분일 때의 표본크기 결정 공식 : $n = \dfrac{N \sum_{h=1}^{H} N_h S_h^2}{N^2 D + \sum_{h=1}^{H} N_h S_h^2}$

(1) 비례배분법

예제 4-7

▶ 비례배분법 적용 예제

층	규모(종업원 수)	제조업체의 수
1	49 인 이하	18,000
2	50 – 99 인	4,000
3	100 – 249 인	2,000
4	250 인 이상	500

층화임의추출법에 의해 500개 표본 추출을 위한 비례배분법

- 층1 : $n_1 = 500 \times \frac{18,000}{24,500} \fallingdotseq 367$
- 층2 : $n_2 = 500 \times \frac{4,000}{24,500} \fallingdotseq 82$
- 층3 : $n_3 = 500 \times \frac{2,000}{24,500} \fallingdotseq 41$
- 층4 : $n_4 = 500 \times \frac{500}{24,500} \fallingdotseq 10$

학/습/목/차

1. 모수추정 – 모총계, 모비율
2. 표본크기의 결정
3. 표본배분법 – 비례배분법
4. 엑셀을 활용한 실습

〈실습하기〉에서 자세히 다룸

이 강의는

강의용 휴대폰(U-KNOU 서비스 휴대폰)으로도

다시 볼 수 있습니다.