

11강 집락추출법(2)

정보통계학과 이기재교수

학/습/목/차

1. 모평균과 모총계 추정 : 집락의 크기가 다른 경우

2. 확률비례추출법의 개념

3. 확률비례추출법에서 추정

4. 엑셀을 활용한 실습

집락추출법 : 집락의 크기가 동일한 경우

- A 개의 집락, 각 집락은 B 개의 기본단위로 구성된 모집단을 대상으로 함
 - ▶ 모집단 내의 전체 조사단위 수 : $N = A \times B$
 - ▶ A 개의 집락 중에서 a 개 집락을 단순임의추출법으로 추출
 - ▶ 추출된 집락 내의 모든 기본단위 B 개를 조사
 - ➔ 조사되는 기본단위 수 : $n = a \times B$
 - 추출률 : $f = a \times B / A \times B = a / A$

추정 : 집락의 크기 동일(2)

- 모평균 μ 에 대한 추정

- ▶ $\bar{y}_{cl} = \sum_{i=1}^a \bar{y}_i / a$

- ▶ 단, $\bar{y}_i = \sum_{j=1}^B y_{ij} / B$: 표본 집락 내의 조사값 평균

- ▶ $\hat{V}(\bar{y}_{cl}) = \left(1 - \frac{a}{A}\right) \frac{s_b^2}{a}$, 여기서 $s_b^2 = \frac{\sum_{i=1}^a (\bar{y}_i - \bar{y}_{cl})^2}{a-1}$

※ 표본 집락 내의 모든 기본단위의 조사값 평균을 집락에 대한 조사값으로 간주하여 계산하기 때문에 단순임의추출법의 분산 추정과 유사함

추정 : 집락의 크기 동일(2)

- 모총계 τ 에 대한 추정

- ▶ 모평균 추정량 \bar{y}_{cl} 에 모집단의 총수(N)를 곱하여 구함

- ▶ $\hat{\tau}_{cl} = N \times \bar{y}_{cl}$

- ▶ $V(\hat{\tau}_{cl}) = N^2 V(\bar{y}_{cl})$

- 모비율 p 에 대한 추정

- ▶ 모평균 추정의 특수한 경우임

집락추출법 : 집락의 크기가 다른 경우

- 현실에서는 집락 크기가 서로 다른 경우가 대부분임
 - ▶ 표본설계 당시에는 같은 크기의 집락이라고 해도 추출틀의 과대 포함이나 누락 등의 이유로 실제의 집락 크기는 서로 다른 경우가 대부분임
 - ▶ 집락의 크기가 동일한 경우는 특수한 경우임

사용 기호 (1)

- ▶ A : 모집단에서 집락의 전체 총수
- ▶ B_i : i 집락에 포함된 기본단위의 수
- ▶ $N = \sum_{i=1}^A B_i$: 모집단에서 단위의 총수
- ▶ $\bar{B} = \sum_{i=1}^A B_i / A = N / A$: 모집단의 평균 집락의 크기
- ▶ a : 표본으로 추출된 집락의 수

사용 기호 (2)

▶ y_{ij} : i 집락 내 j 단위의 조사값

▶ $y_i = \sum_{j=1}^{B_i} y_{ij}$: i 집락 내의 조사값의 합계

▶ $\mu = \sum_{i=1}^A \sum_{j=1}^{B_i} y_{ij} / N$: 모평균

▶ $\tau = \sum_{i=1}^A \sum_{j=1}^{B_i} y_{ij}$: 모총계

모평균의 추정

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

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

➔ 비추정량의 분산식에서 유도됨

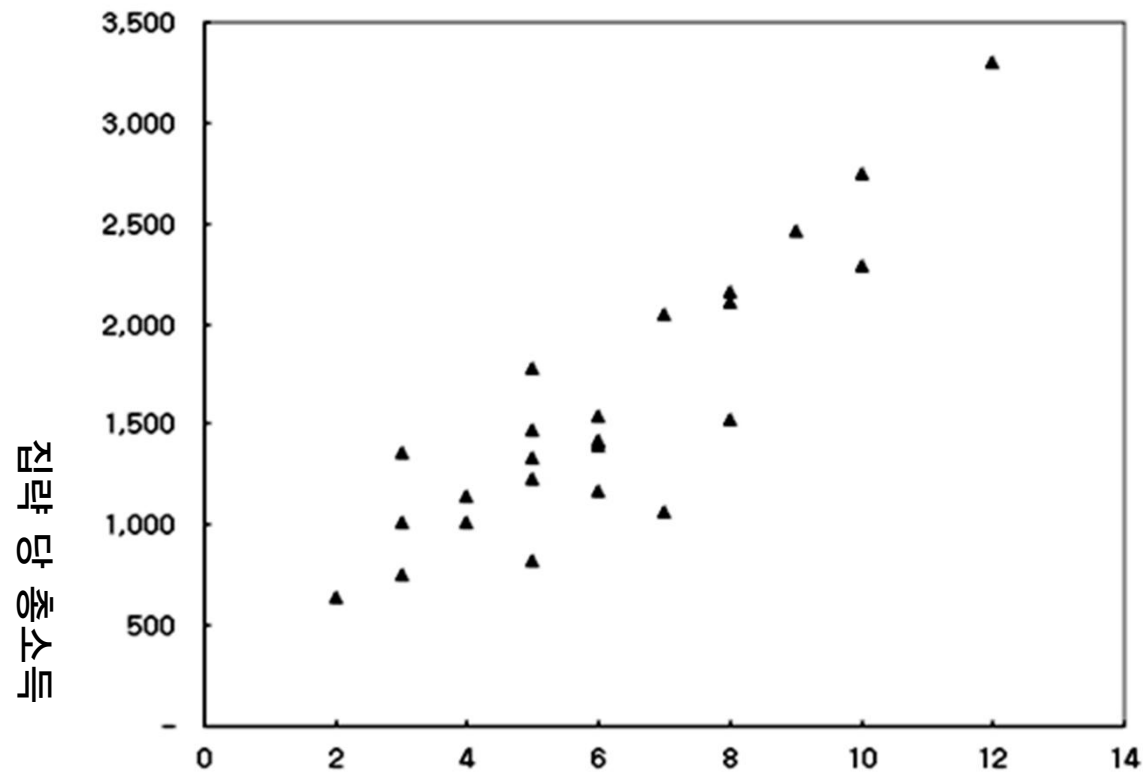
〈참고〉 $r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$, $\hat{V}(r) = \frac{N-n}{N} \frac{1}{n} \frac{1}{\mu_x^2} \frac{\sum_{i=1}^n (y_i - r x_i)^2}{n-1}$

모평균의 추정 예(1)

<표 6-1> 표본 집락에 대한 소득 조사 자료

집락 i	거주자 수 B_i	집락당 총소득 y_i (만원)	집락 i	거주자 수 B_i	집락당 총소득 y_i (만원)
1	8	2,160	14	10	2,750
2	12	3,300	15	9	2,460
3	4	1,140	16	3	1,010
4	5	820	17	6	1,420
5	6	1,410	18	5	1,330
6	6	1,170	19	5	1,230
7	7	2,050	20	4	1,010
8	5	1,780	21	6	1,390
9	8	1,520	22	10	2,290
10	3	1,360	23	7	1,060
11	2	640	24	3	750
12	6	1,540	25	8	2,110
13	5	1,470	계 $\sum_{i=1}^{25} B_i = 153$ $\sum_{i=1}^{25} y_i = 39,150$		

모평균의 추정 예(2)



각 집락의 거주자 수

$$\bar{b} = \frac{\sum_{i=1}^{25} B_i}{25} = \frac{153}{25} = 6.12, \quad cv(\bar{b}) = \frac{\sqrt{\hat{V}(\bar{b})}}{\bar{b}} = \frac{0.23218}{6.12} = 0.079$$

모총계의 추정(1)

- N 을 알 수 있는 경우

$$\blacktriangleright \hat{\tau}_{cl} = N\bar{y}_{cl} = N \times \frac{\sum_{i=1}^a y_i}{\sum_{i=1}^a B_i}$$

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

모총계의 추정(2)

- N 을 알 수 없는 경우

- ▶ $\hat{\tau} = A \bar{y}_t = A \times \frac{\sum_{i=1}^a y_i}{a}$

- ▶ 단, \bar{y}_t : a 개 표본집락에 대한 집락 총계의 평균

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

학/습/목/차

1. 모평균과 모총계 추정 : 집락의 크기가 다른 경우

2. 확률비례추출법의 개념

3. 확률비례추출법에서 추정

4. 엑셀을 활용한 실습

확률비례추출법

모집단을 구성하는 집락들을 같은 확률로 추출하지 않고
집락의 규모측도에 비례하여 추출하는 방법

- ▶ Probability Proportional to Size Sampling 또는
PPS Sampling이라고 함
- ▶ 각 집락을 집락의 크기에 비례하여 뽑는 방법이
일반적으로 사용되는 방법임

규모측도(measure of size : MOS)

- ▶ 조사항목과 상관관계가 높은 것을 채택하면 추정의 효율이 높아짐
- ▶ 흔히 규모측도로 사용되는 것은 본 조사 전에 실시된 센서스 결과

표본추출법

- ▶ B_i : i 번째 집락의 크기,

$$y_i = \sum_{j=1}^{B_i} y_{ij} : \text{집락 내 조사값의 합계}$$

- ▶ 사전에 알려진 집락의 크기 B_i 에 비례하여
복원확률비례추출
 - ➔ 누적총계법 이용
 - ➔ 추출방법이 달라짐에 따라 추정량도 달라짐

복원확률비례추출법의 예 : 누적총계법

✚ H 백화점 A=8 개의 분점 중에서 a=3 분점을 점원수에 비례하여 추출

- 모든 분점에 대해서 집락 크기의 누적규모 계산

분 점	점 원 수 [B_i]	누 적 범 위
1	1,200	1 ~ 1,200
2	450	1,201 ~ 1,650
3	2,100	1,651 ~ 3,750
4	860	3,751 ~ 4,610
5	2,850	4,611 ~ 7,460
6	1,950	7,461 ~ 9,410
7	390	9,411 ~ 9,800
8	3,200	9,801 ~ 13,000
합 계	13,000	

- 00001에서 13,000사이에서 3개의 난수 선정

학/습/목/차

1. 모평균과 모총계 추정 : 집락의 크기가 다른 경우

2. 확률비례추출법의 개념

3. 확률비례추출법에서 추정

4. 엑셀을 활용한 실습

모총계 추정 (1)

- A 개 집락의 모집단에서 a 개 표본 집락을 확률비례추출하는 경우

- 모총계 추정

- ▶
$$\hat{\tau}_{PPS} = \frac{N}{a} \sum_{i=1}^a \frac{y_i}{B_i} = \frac{N}{a} \sum_{i=1}^a \bar{y}_i$$

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

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

확률비례추출법 추정 (2)

- 기본단위당 모평균

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

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

학/습/목/차

1. 모평균과 모총계 추정 : 집락의 크기가 다른 경우

2. 확률비례추출법의 개념

3. 확률비례추출법에서 추정

4. 엑셀을 활용한 실습

↳ <실습하기>에서 자세히 다룸



Korea National Open University
이 강의는
강의용 휴대폰(U-KNOU 서비스 휴대폰)으로도
다시 볼 수 있습니다.

다시 볼 수 있습니다.