

〈권이태〉

Problem. 1. 한 도시의 소득에 관해 추정하고자 한다. 추출틀로 사용할 성인 거주자들에 대한 목록이 없어 도시를 415개의 직사각형 블록으로 나눈 뒤, 25개의 블록을 임의로 조사하였다. 다음은 추출된 25개 집락에 대해 집락당 거주자 수와 총소득을 조사한 결과이다. a 는 정해진 상수이다.

	평균	표준편차
거주자 수(m)	6.040	2.371
집락 총 소득(y)	53.160	21784
$y - am$	0	25189

(a) 도시에 총 2500명의 거주자가 있을 때, 거주자들의 평균 수입을 추정하여라.

(b) 평균소득을 추정오차의 한계가 400이 되도록 추정하고자 할 때, 필요한 집락의 크기를 구하여라.

(c) 총 거주자의 수를 모른다고 가정하자. 거주자들의 평균 수입을 추정하여라. a 와는 어떤 관련이 있는가?

(d) 총 거주자의 수를 모른다고 가정하자. 평균소득을 추정오차의 한계가 400이 되도록 추정하고자 할 때, 필요한 집락의 크기를 구하여라.

Problem. 2. 층화집락추출을 고려하자. 우리는 각 원소들에 대한 모평균을 추정하는 것을 목표로 하고 있다. L 개의 층이 존재하고 i 번째 층에는 N_i 개의 집락이 존재함이 알려져 있다. i 번째 층 j 번째 집락의 k 번째 개체가 가진 관심변수가 y_{ijk} 이며, i 번째 층 j 번째 집락의 개체수가 m_{ij} 이고, i 번째 층에서 n_i 개의 집락 $\{a_{i1}, a_{i2}, \dots, a_{in_i}\}$ 을 추출하였다고 하자.

(a) $M = \sum_{i=1}^L \sum_{j=1}^{N_i} m_{ij}$ 를 모를 때, 모평균의 추정량과 그 분산추정량을 제시하시오.

(b) $M = \sum_{i=1}^L \sum_{j=1}^{N_i} m_{ij}$ 를 알 때, 모평균의 불편추정량과 그 분산추정량을 제시하시오.

Problem. 3. 39800개의 기업고객정보를 3980개의 폴더에 10개씩 정리하여 저장하고 있는 은행에서, 대출을 허용한 고객기업의 비율 p 를 추정하려 한다. 이를 위하여 $n = 40$ 개의 폴더를 단순임의추출하여 표본폴더 i 에서 대출이 승인된 기업수 y_i 를 조사하여 다음과 같은 결과를 얻었다.

$$\sum_{i=1}^n y_i = 185, \quad \sum_{i=1}^n y_i^2 = 1263$$

(a) p 에 관한 불편추정량을 구하여라.

(b) 앞에서 구한 추정량의 분산을 추정하여라.

Problem. 4. (a) 모집단 $U = \{1, 2, 3, 4, 5\}$ 이 5개의 원소를 가지고 있으며 각각은 y_i 값으로 5, 10, 6, 0, 0을 가진다. 이때 표본분포가

$$P(\{1, 3, 5\}) = \frac{4}{10}, P(\{2, 3, 5\}) = \frac{3}{10}, P(\{1, 2, 4\}) = \frac{1}{10}, P(\{3, 4, 5\}) = \frac{2}{10}$$

이라고 한다. 만약 $s = \{1, 2, 4\}$ 가 sample로 뽑힌다면 이때의 모집단 총계의 추정량과 그 추정분산을 구하시오.

(b) 확률변수 X_1, X_2, \dots, X_n 이 서로 독립이고 $\mathbb{E}[X_i] = \mu$, $\mathbb{V}(X_i) = \sigma_i^2$ 이 성립한다면, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 의 비편향 분산 추정량

$$\hat{\mathbb{V}}(\bar{X}) = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$

은 비편향인 동시에 항상 음이 아님을 보여라.

- (c) 비복원 단순임의추출을 통해 모평균을 추정하려 한다. 이때 전체 모집단의 개체수는 1000, 표본수는 100이다. i 번째 표본의 관심변수는 y_i 라고 하자.

$$\sum_{i=1}^{100} y_i = 400, \quad \sum_{i=1}^{100} y_i^2 = 3600$$

임이 알려져 있을 때, 모평균의 추정량과 그 분산의 추정량을 구하여라. 또한 (b)에서 X_i 자리에 y_i 를 넣어 만든 분산추정량을 계산하여라.

- (d) (c)에서, 두 분산추정량의 차이가 무엇으로부터 비롯되는지 밝히시오. 만약 두 추정량 중 불편추정량이 아닌 것이 있다면, 해당 방법으로 추정량이 불편추정량이 아님에도 사용될 가치가 있는지 (a)를 참고하여 논하시오.

Problem. 5. 아래와 같은 이단계 집락추출을 생각하자.

1. 제 1단계 추출: n_I 개의 집락을 M_i 에 비례하게 PPS로 추출한다.
 2. 제 2단계 추출: 각 표본집락에서 m 개의 원소를 SRS로 추출한다.
- (a) 이러한 이단계 집락추출로부터 얻은 모집단 평균의 추정량과 그 분산추정량을 구하여라. 표본에서 얻은 i 번째 집락의 j 번째 원소가 가진 관심변수를 y_{ij} 라고 하자.

- (b) 한국동의 가구당 인구수를 추정하고자 한다. 이를 위하여 제 1단계 추출에서는 $n_I = 3$ 개의 표본집락을 집락 내의 가구수 M_i 에 비례하게 PPS 추출을 사용하여 추출하였다. 아래는 실제 표본집락별 분포이다.

표본집락 (i)	집락 내 가구수(M_i)	집락 내 가구당 인원수 평균값(\bar{t}_i)
1	20	5.25
2	24	5.50
3	24	4.50
4	18	5.00

실제 전체 인구수를 구하시오.

(c) 그 후 제 2단계 추출에서 각 추출된 표본집락내의 M_i 가구 중 $m_i = 4$ 가구를 비복원 단순임의추출로 추출하였다. 그 결과가 아래와 같다.

표본집락 (i)	표본가구 (j)	총 인원수(t_{ij})	6세 이하 인원수 y_{ij}
1	1	8	2
1	2	7	2
1	3	7	1
1	4	6	1
2	1	8	0
2	2	12	1
2	3	10	3
2	4	11	1
3	1	4	2
3	2	5	3
3	3	5	2
3	4	6	1

한국동에서 6세 이하의 전체 인구수 비율의 추정치를 구하고, 그 분산 추정치를 구하시오.

(d) 분산추정치를 통해 표본설계효과를 추정하시오.