

표본론 2023년 중간시험 (월, 수 9:00~10:15)

[1] (20점) 모집단이 10개의 원소 {11, 13, 14, 16, 17, 20, 21, 23, 24, 25}로 이루어져 있다고 하자.

[1-1] 5개를 단순확률추출(simple random sampling)한 표본이 {11, 13, 17, 23, 25}이라고 하자. 모평균을 추정하고 오차의 한계(bound on the error of estimation)를 구하시오. 이때 모집단의 크기가 10인 것은 알고 있으나 모집단 원소 10개의 값은 모른다고 가정하시오.

[1-2] 5개를 단순확률추출해서 얻은 표본평균을 \bar{y} 라고 할 때 \bar{y} 의 기댓값 $E(\bar{y})$ 와 참분산 $V(\bar{y})$ 를 각각 구하시오.

[2] (30점) 쿠바에서 미국 플로리다로 이주한 사람들을 대상으로 1990년대 후반에 조사를 실시하였다. 1001명에 대해 응답을 얻었는데 응답자 중에서 백인은 87%, 흑인은 13%이었다. 쿠바 인구 센서스에 따르면 쿠바 인구 중에서 60%가 백인, 40%가 흑인이라고 한다. “쿠바 혁명이 교육에서 성과를 이루었다고 생각하는가?”라는 질문에 대해 아래 표와 같은 결과를 얻었을 때 다음 질문에 답하시오.

	Yes	No	Total
White	566	304	870
Black	102	29	131
Total	668	333	1001

[2-1] 1001명의 응답자를 전체 이민자의 단순확률표본(a simple random sample)으로 간주하고 “예”라고 답한 사람의 모비율을 추정하시오. 오차의 한계도 구하시오.

[2-2] “예”라고 답한 사람의 모비율에 대한 사후층화(poststratification) 추정량을 구하시오. 오차의 한계도 구하시오.

[2-3] 사후층화 추정량이 더 나은 추정량이라고 할 수 있는지에 대해 논하시오.

[3] (20점) 층화확률추출(stratified random sampling)이 언제 단순확률추출(simple random sampling)보다 정밀도(precision)를 더 높일 수 있는지에 대해 수식을 이용하여 설명하시오. 이때 두 추출방법의 조사비용은 동일하다고 가정하시오.

[4] (30점) 크기 N 인 모집단을 L 개의 층(strata)으로 나누어 층화추출(stratified random sampling)을 하려고 한다. N_i 는 i 번째 층의 크기를 나타내며 $N = N_1 + N_2 + \dots + N_L$ 이다. 각 층에서 추출하는 표본의 크기 n_i 를 층의 크기 N_i 에 비례해서 할당(proportional allocation)했을 때,

[4-1] i 번째 층에 있는 임의의 한 원소가 추출될 확률이 단순확률추출(simple random sampling) 했을 때의 확률과 같음을 증명하시오.

[4-2] 모평균에 대한 추정량의 형태가, 마치 단순확률추출을 한 것으로 간주하고 구한 추정량과 같아짐을 증명하시오. 즉, $\sum_{i=1}^L N_i \bar{y}_i / N = \sum_{i=1}^L \sum_{j=1}^{n_i} y_{ij} / n$ 임을 증명하시오. 단, y_{ij} 는 i 층에서 추출한 j 번째 관측값이고, $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$ 이다.

[4-3] 그렇다면 ‘비례배분에 의한 층화추출’과 ‘단순확률추출’은 아무런 차이가 없는 것인가?