

표본론 2021년 중간시험

[1] 19대 대통령선거 결과 예측을 위해 2017년 5월 1일 한국갤럽조사연구소에서 1015명을 대상으로 실시한 전화조사에서 자유한국당 홍준표 후보 지지율이 16%, 국민의당 안철수 후보 지지율이 20%이었다. 만약 1015명을 단순확률표본(simple random sample)이라고 가정할 수 있다면, 조사 당시 두 후보의 지지율에 유의한(significant) 차이가 있다고 할 수 있는가?

[2] 250 그루의 나무 중에서 6 그루를 단순확률추출해서 밀동면적 x 와 부피 y 를 측정하였다. 250 그루 나무 밀동면적의 총합 τ_x 가 75인 것을 알고 있다. 250 그루 나무의 총부피 τ_y 를

[2-1] y 자료만으로 추정시오. (τ_y 에 대한 신뢰구간도 구해야 함)

[2-2] x 자료를 같이 이용하여 추정하시오. (τ_y 에 대한 신뢰구간도 구해야 함)

나무 번호	x	y	나무 번호	x	y
1	0.3	6	4	0.9	19
2	0.5	9	5	0.7	15
3	0.4	7	6	0.2	5

[3] 3개의 층(strata)으로 이루어진 모집단의 특성이 아래 표와 같다고 할 때

[3-1] 모평균(population mean)과 모분산(population variance)을 구하시오.

[3-2] 이 모집단과 특성이 비슷한 모집단에서 표본크기가 50인 표본을 추출하려고 한다. 네이만 배분(Neyman allocation) 방법을 적용하기로 할 때 각 층에서 몇 개의 표본을 추출해야 하는지 정하시오.

층 번호	층 크기	층 평균	층 분산
1	110	7.5	1.30
2	90	7.0	2.05
3	70	11.0	1.10

[4] “표본을 추출할 때 표집단위(sampling unit)가 추출될 확률이 모두 같아지도록 해야 한다”는 주장에 대해 논하시오.

[5] 선거예측조사는 전화조사로 실시하는데 자동응답시스템(ARS) 방식인 경우 응답률이 5% 내외이고, 면접방식인 경우 응답률이 15% 내외이다. (이 때 응답률은 결번이거나 전화를 받지 않은 경우를 제외하고 전화연결이 된 통화 중에서 응답을 얻은 비율을 의미한다.) 선거예측조사에서 “ARS 방식과 같이 응답률이 5% 정도로 낮은 조사는 믿을 수가 없다”는 주장에 대해 논하시오.

표본론 2021년 기말시험

- [1] (20점) 묘목장에 있는 묘목의 평균 키를 추정하기 위해 묘목장을 조금씩 다른 크기를 갖는 5개의 구역으로 나누었다. 랜덤하게 추출한 5개의 구역에서 10%씩 묘목을 추출하여 아래와 같은 자료를 얻었다. 묘목장에 있는 묘목의 평균 키를 추정하고 오차의 한계를 구하시오.

구역	구역에 있는 묘목의 수	추출한 묘목의 수	묘목의 키
1	52	5	36, 33, 36, 30, 39
2	56	6	30, 27, 21, 27, 24, 30
3	60	6	18, 15, 21, 15, 18, 12
4	46	5	21, 24, 21, 21, 18
5	49	5	30, 33, 39, 36, 36

- [2] (20점) 산업공단에 인접한 어떤 지역에서 대기 중에 있는 특정한 독성 입자의 밀도를 추정하고자 한다. 100 세제곱센티미터 단위의 대기 표본을 추출하는데, 추출한 대기 표본에 있는 독성 입자의 수를 세기는 어렵지만, 독성 입자의 유무를 판단하기는 쉽다고 한다. 500개의 대기 표본을 추출하여 관측한 결과 280개에서 독성 입자가 발견되었다고 할 때, 대기 중 특정 독성 입자의 밀도(단위: 개/cm³)를 추정하고 오차의 한계를 구하시오.
- [3] (20점) 정부의 주택정책에 대한 의견을 묻기 위해 전화조사를 실시했다. 모집단은 특정 지역에 거주하는 성인이다. 200명이 응답하였고 응답률은 10%이었다. 응답자의 성별과 거주하는 주택의 소유 여부에 대해 살펴보았더니 다음 표와 같았다.
- [3-1] 만약 모집단의 크기가 10만명이고, 성별 비율이 1:1, 자가와 전월세 비율도 1:1이라고 알려져 있다면, 성별과 주택 소유 여부로 정해지는 네 개의 층의 크기를 갈퀴질(raking) 또는 림 가중(RIM weighting) 방식으로 추정하시오. 각 칸의 도수를 계산하는 과정에서 소수 첫째 자리에서 반올림해 칸 도수가 항상 정수가 되도록 하고, 행의 합이나 열의 합이 목표한 도수 100에서 1을 초과해 벗어나지 않는 경우는 계산을 중단해도 좋다. 예를 들어 두 행과 두 열의 합이 각각 100이 되어야 하는데 99에서 101 사이라면 수렴했다고 간주하고 더 이상 계산을 하지 않아도 좋다.
- [3-2] “위 방식으로 추정된 층의 크기는 실제 층 크기의 비편향추정량(unbiased estimator)이 된다”는 주장의 타당성에 대해 논하시오.

	자가	전월세	합
남성	28	72	100
여성	42	58	100
합	70	130	200

- [4] (40점) 모집단이 세 개의 원소 {1, 2, 8}로 구성되어 있다고 가정하자. 여기서 두 개의 원소를 추출하여 모집단의 총합을 추정하려고 한다. 총합의 참값은 $\tau = 11$ 이다. 표본크기 2인 표본을 복원추출(sampling with replacement) 하였으며, 모집단의 세 원소를 추출할 확률을 각각 0.1, 0.1, 0.8로 하였다.
- [4-1] 아래 표의 빈 칸을 모두 채우시오. (빈 칸의 수는 15개이다.)
- [4-2] 이 예에서 총합 추정량 $\hat{\tau}$ 가 τ 의 비편향추정량(unbiased estimator)임을 표에서 구한 수치로 증명하시오.
- [4-3] 총합 추정량 $\hat{\tau}$ 의 모분산(population variance), $V(\hat{\tau})$ 을 구하시오.
- [4-4] 만약 표본크기 두 개를 비복원으로 단순확률추출(simple random sampling) 했다면 총합 추정량 $\hat{\tau}_{srs}$ 의 모분산(population variance) $V(\hat{\tau}_{srs})$ 은 얼마인가? 모집단 크기가 3인 것은 알고 있다고 가정하시오.

표본	표본 추출 확률	$\hat{\tau}$	$\hat{V}(\hat{\tau})$
(1, 2)	0.02	15	25
(1, 8)			
(2, 8)			
(1, 1)			
(2, 2)			
(8, 8)			

[1] 19대 대통령선거 결과 예측을 위해 2017년 5월 1일 한국갤럽조사연구소에서 1015명을 대상으로 실시한 전화조사에서 자유한국당 홍준표 후보 지지율이 16%, 국민의당 안철수 후보 지지율이 20%이었다. 만약 1015명을 단순확률표본(simple random sample)이라고 가정할 수 있다면, 조사 당시 두 후보의 지지율에 유의한(significant) 차이가 있다고 할 수 있는가?

[2] “표본을 추출할 때 표집단위(sampling unit)가 추출될 확률이 모두 같아지도록 해야 한다”는 주장에 대해 논하시오.

[3] 어떤 TV 토론 프로그램에서 시청자들에게 전화로 토론 주제에 대한 찬반조사를 실시하였다. 조사에 참여한 사람 250명 중에서 여성이 30%, 남성이 70%이었고 여성의 찬성비율은 65%, 남성의 찬성비율은 40%이었다. 토론 프로그램의 사회자가 얘기하기를 “인구 전체 남성과 여성 비율이 50%이므로 이에 맞게 가중값을 부여해

$$0.5 \times 0.65 + 0.5 \times 0.4 = 0.525$$

와 같이 계산해서 국민 전체의 52.5%가 찬성입니다”라고 했다. 이런 식으로 가중값을 부여해 찬성비율을 조정하는 것에 대해 어떻게 생각하는가?

[4] 변수 y 에 대한 모집단 총합을 추정하고자 할 때, y 의 관측값만 이용하여 추정하지 않고 또 다른 변수 x 를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건(들)이 만족될 것 같은 구체적인 예를 들어 보시오.

[5] 낮은 응답률을 보이는 전화여론조사의 신뢰도에 관한 문제이다. 중앙선거여론조사심의위원회에 등록된 전국 단위 대선 및 정당 지지율 여론조사 66개를 분석한 결과 응답률 최저는 3.1%, 최고는 23.9%였으며 평균은 약 9.2%였다. 낮은 응답률을 보인 조사에 대한 다음 두 사람의 주장에 대해 논하시오.

A: “1천 명을 뽑아내기 위해 1만 명에게 물었는지, 2만 명에게 물었는지는 신뢰도와는 직결되지 않는 문제이다”

B: “여론조사를 하면 90% 이상이 응답을 하지 않는다. 지금 여론조사라는 것은 경향성만 보는 것이지 큰 의미가 없다”