

A. 크기가 $N = 5$ 인 다음의 모집단을 고려하자.

k	1	2	3	4	5
y_k	6	8	9	10	12
z_k	1	0	0	1	1

아래와 같이 크기가 $n = 3$ 인 가능한 모든 표본과 대응하는 표본설계가 다음과 같이 주어졌음을 가정하자.

Sample	$P(S)$	Sample	$P(S)$
$S_1=(1, 2, 3)$	0.06	$S_6=(1, 4, 5)$	0.10
$S_2=(1, 2, 4)$	0.07	$S_7=(2, 3, 4)$	0.11
$S_3=(1, 2, 5)$	0.08	$S_8=(2, 3, 5)$	0.12
$S_4=(1, 3, 4)$	0.09	$S_9=(2, 4, 5)$	0.13
$S_5=(1, 3, 5)$	0.10	$S_{10}=(3, 4, 5)$	0.14

- (a) 개체표본포함확률 $\pi_i, i = 1, \dots, 5$ 을 계산하시오.
- (b) 주어진 표본설계 아래서 표본평균 $\bar{y} = 3^{-1} \sum_{k \in S} y_k$ 과 표본비율 $\hat{p} = 3^{-1} \sum_{k \in S} z_k$ 의 표본추출분포(sampling distribution)를 구하시오.
- (c) 주어진 표본설계 아래서 비편향 추정량 $\bar{y}_{unb} = 5^{-1} \sum_{k \in S} (y_k / \pi_k)$ 과 $\hat{p}_{unb} = 5^{-1} \sum_{k \in S} (z_k / \pi_k)$ 의 표본추출분포(sampling distribution)을 구하시오.
- (d) (b)와 (c)에 주어진 추정량 중 어느 추정량의 사용을 권유하며 그 이유는 무엇입니까?

B. SAS의 surveyselect를 이용하여 크기가 $n = 100$ 인 표본을 비복원 단순임의추출법을 통해 추출하고 이를 이용하여 아래의 문항에 답하시오. 표본추출을 위해서는 본인의 학번을 SEED번호로 지정하시오.(수업시간에 고려한 agpop.dat과 첨부된 프로그램을 이용하여 추출틀을 구성하시오.)

- (a) 표본추출과정을 요약한 SAS output을 제시하고 아래의 문항에 답하시오.
- 표본 추출 비율(sampling fraction), f 는?
 - 특정한 표본이 추출될 확률은?
 - 특정한 카운티(county)가 추출될 확률은?
- (b) 각 3개 년도 별 농장으로 사용되는 카운티 별 평균면적(mean number of acres per county devoted to farms)을 추정하고 추정량의 표준오차(standard errors of the mean estimator)를 제시하시오.

(c) SURVEYMEANS 프로시저를 이용하여 규모가 큰 농장의 수가 규모가 작은 농장의 수 보다 많은 카운티의 비율을 추정하고 추정량의 표준오차를 구하시오.

C. 첨부된 **golfsrs.dat**는 미국 내 16,883 골프장 중 단순임의표본추출법을 통해 추출된 120개의 골프장을 포함하고 있다(참고: www.golfcourse.com). 아래의 명령문을 이용하여 자료를 읽고 아래의 문제에 답하시오. 자료에 대한 설명은 첨부된 **golfsrs.pdf**를 참고하시오.

```
INPUT rn state $ course $ holes type $ yearblt wkday18
      wkday9 wkend18 wkend9 backtee rating par cart18
      cart9 caddy $ pro $ ;
```

(a) 골프 프로페셔널이 이용 가능한 코스를 **보유한** 골프장에서 9홀을 경기하기 위한 주중 평균 그린피(mean weekday greens fee to play 9 holes)를 추정하고 이의 표본오차(standard error)를 구하시오.

(b) 골프 프로페셔널이 이용 가능한 코스를 **보유하지 않은** 골프장에서 9홀을 경기하기 위한 주중 평균 그린피(mean weekday greens fee to play 9 holes)를 추정하고 이의 표본오차(standard error)를 구하시오.

D. 비복원 단순임의추출 하에서 모평균 \bar{y}_U 의 신뢰수준 $100 \times (1 - \alpha)\%$ 의 신뢰구간은 $\bar{y}_{unb} (1 \pm A)$ 로 표현할 수 있다. 여기서

$$A = z_{1-\alpha/2} cv_{ys} \sqrt{\frac{1-f}{n}},$$

$cv_{ys} = S_{ys}/\bar{y}_{unb}$ 과 $f = n/N$ 을 각 각 나타낸다. $cv_{ys} = cv_{yU} \approx S_{yU}/\bar{y}_U$ 과 $f \approx 0$ 의 가정 하에서 $A \leq 3\%$ 와 다음의 주어진 cv_{yU} 을 만족하기 위한 표본크기를 산출하시오. $\alpha = 0.05$ 이다.

(a) $cv_{yU} = 0.5$

(b) $cv_{yU} = 1.0$

(c) $cv_{yU} = 1.5$