

---

## STAT 521: Homework Assignment 1

Due in class on February 13, 2024

### Problem 1: (15 pt)

Consider the following sampling design from a finite population  $U = \{1, 2, 3\}$ . Let  $y_i$  be the study item of interest in unit  $i$  in the population. We are interested in estimating the population total of  $y$ .

Sample (A)	$Pr(A)$	HT estimator	HT var. est.	SYG var. est.
$A_1 = \{1, 2\}$	0.5			
$A_2 = \{1, 3\}$	0.25			
$A_3 = \{2, 3\}$	0.25			

1. Compute the HT estimators and the two variance estimators for each sample. Check the unbiasedness of the variance estimators. (May assume  $y_1 = 16, y_2 = 21, y_3 = 18$  here only.)
2. Now, consider the special case of  $y_k = \pi_k$ , where  $\pi_k$  is the first order inclusion probability of unit  $k$ . What is the variance of the HT estimator?
3. Also, under the case of  $y_k = \pi_k$ , compute HT variance estimator and SYG variance estimator for each sample. (They are not the same.) Which variance estimator do you prefer? Why ?

---

**Problem 2:** (15pt)

Let  $U$  be a finite population of size  $N$ . We define the following sampling design: we first select a sample  $A_1$  according to a simple random sampling (without replacement) of fixed size  $n_1$ . We then select a sample  $A_2$  in  $U$  outside of  $A_1$  according to a simple random sampling design without replacement of fixed size  $n_2$ . The final sample  $A$  consists of  $A_1$  and  $A_2$ .

1. What is the sampling distribution of  $A$ ? What is interesting about this result?
2. We define the estimator of  $\bar{Y}$ , the finite population mean of  $y$ , by

$$\bar{y}_\alpha = \alpha \bar{y}_1 + (1 - \alpha) \bar{y}_2$$

with  $0 < \alpha < 1$ , where  $\bar{y}_1$  is the sample mean of  $y$  in  $A_1$  and  $\bar{y}_2$  is the sample mean of  $y$  in  $A_2$ . Show that  $\bar{y}_\alpha$  is unbiased for  $\bar{Y}$  for any  $\alpha$ .

3. Find the optimal value of  $\alpha$  that minimizes the variance of  $\bar{y}_\alpha$ .

[Hints for (3):] Since

$$V(\bar{y}_\alpha) = \alpha^2 V(\bar{y}_1) + (1 - \alpha)^2 V(\bar{y}_2) + 2\alpha(1 - \alpha) \text{Cov}(\bar{y}_1, \bar{y}_2),$$

it is minimized at

$$\alpha^* = \frac{V(\bar{y}_2) - \text{Cov}(\bar{y}_1, \bar{y}_2)}{V(\bar{y}_1) + V(\bar{y}_2) - 2\text{Cov}(\bar{y}_1, \bar{y}_2)}.$$

---

**Problem 3:** (10 pt)

A community in the San Francisco Bay area consists of approximately 100,000 persons. It is desired to estimate in this community, the proportion of persons who are not covered by some form of health insurance. One would like to be 95% certain that this estimate is within 15% of the true proportion, which is believed to lie somewhere between 10% and 20% of the total population. That is, we wish to achieve

$$P \left( \frac{|\hat{P} - P|}{P} \leq 0.15 \right) = 0.95$$

where  $P$  is the true proportion satisfying  $0.1 \leq P \leq 0.2$ . Assuming simple random sampling, how large a sample is needed?