## Cluster sampling with estimation

Suppose a population of size $N$ is divided into $K = N/M$ groups of size $M$. We select a sample of size $n = km$ the following way:

- First we select $k$ groups out of $K$ groups by simple random sampling.
- We then select $m$ units in each group selected on the first step by simple random sampling.
- The estimate of the population mean is the average $\bar{Y}$ of the sample.

Let $\mu_i$ be the population average in the $i$-th group for $i = 1, 2, \ldots, K$, and let $\sigma_i^2$ be the population variance in the $i$-th group for $i = 1, 2, \ldots, K$.

a. Show that we can write the estimator as

$$\bar{Y} = \frac{1}{k} \sum_{i=1}^{K} \bar{Y}_i I_i \, ,$$

where

$$I_i = \begin{cases} 1 & \text{if the } i\text{-th group is selected.} \\ 0 & \text{otherwise} \end{cases}$$

and $\bar{Y}_i$ is the sample average in the $i$-th group for $i = 1, 2, \ldots, K$. Argue that it is reasonable to assume that the random variables $\bar{Y}_1, \ldots, \bar{Y}_K$ are independent and independent from $I_1, \ldots, I_K$. Show that $\bar{Y}$ is an unbiased estimator of the population mean $\mu$ and show that the variance of $\bar{Y}$ is

$$\mathrm{var}(\bar{Y}) = \frac{M - m}{k(M-1)m} \cdot \frac{1}{K} \sum_{i=1}^{K} \sigma_i^2 + \frac{K - k}{k(K-1)} \cdot \frac{1}{K} \sum_{i=1}^{K} (\mu_i - \mu)^2 \, .$$

b. Suggest an estimate for the quantity

$$\sigma_b^2 = \frac{1}{K} \sum_{i=1}^{K} (\mu_k - \mu)^2 = \frac{1}{K} \sum_{i=1}^{K} \mu_k^2 - \mu^2 \, .$$

Is your estimate unbiased? Can you modify it to be an unbiased estimate?