Final-Assignment-Q.1 (12 marks)

In a list of N individuals, we are interested in a variable y. The individuals are identified by their order on the list, so their order goes from 1 (for the first) to N (for the last). We use systematic sampling with interval k to select n individuals from the list. We assume that: $k = \frac{N}{n} \in \mathbf{N}$.

Let $S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$ be the sample variance for the i^{th} cluster, where $\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$ denote the i^{th} cluster mean.

(a) (4 marks) Show that everything happens as if we selected a unique cluster of individuals from a population pre-divided into clusters. We will specify what the clusters are, what their size is, and how many there are in the population.

(b) (4 marks) Let y_{ij} be the value of y for j^{th} record counted in cluster i, and μ denote the population mean for y.

(i) What is the unbiased estimator $\widehat{\mu}$ of μ? Show that $\widehat{\mu}$ is effectively unbiased.

(ii) What is the expression of the true variance of $\widehat{\mu}$, as a function of μ, the i^{th} cluster mean $\bar{y}_i$ and k.

(c) (2 marks) Considering the natural splitting of the population into k clusters, show that the general expression of the population variance σ_y^2 can be decomposed by

$$\sigma_y^2 = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_i - \mu)^2 + \frac{n-1}{n \times k} \sum_{i=1}^{k} S_i^2$$

(d) (1 mark) Show that if N is large, and if we denote:

$$\text{MSW} = \frac{\text{SSW}}{k(n-1)} = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$$

then we have:

$$\text{Var}(\widehat{\mu}) = \sigma_y^2 - \frac{k(n-1)}{N} \text{MSW}$$

(e) (1 mark) Show that systematic sampling is more precise than simple random sampling if and only if: $\sigma_y^2 < \text{SSW}$ by considering N as very large with respect to n.