

1 Variance estimation for multi-stage sampling

Let U be a finite population and $\mathcal{P}(U)$ the power set of U . Let $p : \mathcal{S} \rightarrow (0, 1]$ be a r -stage sampling design ($r \geq 2$), where $\mathcal{S} \subset \mathcal{P}(U)$ is the set of all admissible samples under the design p . We express the hierarchical structure of the population U , with respect to the r -stage design p , as follows:

$$U = \bigsqcup_{i \in U^{(1)}} U_i^{(2+)} = \bigsqcup_{i \in U^{(1)}} \bigsqcup_{a \in U_i^{(2)}} U_{ia}^{(3)} = \cdots = \bigsqcup_{i \in U^{(1)}} \cdots \bigsqcup_{b \in U_{i \dots}^{(r-1)}} U_{i \dots b}^{(r)} \quad (1.1)$$

where $U^{(1)}$ is the set of all primary sampling units (PSU), and for each PSU $i \in U^{(1)}$, $U_i^{(2)}$ denotes the set of all secondary sampling units (SSU) contained in $i \in U^{(1)}$, and for each SSU $a \in U_i^{(2)}$, $U_{ia}^{(3)}$ denotes the set of all tertiary sampling units (TSU) contained in $a \in U_i^{(2)}$, and so on. Similarly, we express the hierarchical structure of every admissible sample $s \in \mathcal{S}$ as follows:

$$s = \bigsqcup_{i \in s^{(1)}} s_i^{(2+)} = \bigsqcup_{i \in s^{(1)}} \bigsqcup_{a \in s_i^{(2)}} s_{ia}^{(3)} = \cdots \quad (1.2)$$

Let $y : U \rightarrow \mathbb{R}$ be a population characteristic. Let T be its population total, i.e.

$$T = \sum_{u \in U} y_u = \sum_{i \in U^{(1)}} T_i = \sum_{i \in U^{(1)}} \sum_{a \in U_{ia}^{(2)}} T_{ia} = \cdots = \sum_{i \in U^{(1)}} \cdots \sum_{u \in U_{i \dots}^{(r)}} y_u \quad (1.3)$$

Theorem 1.1

If $\hat{T}_i : \mathcal{S}_i^{(2+)} \rightarrow \mathbb{R}$ is an unbiased estimator for T_i , for each PSU $i \in U^{(1)}$, then the random variable $\hat{T} : \mathcal{S} \rightarrow \mathbb{R}$ defined as follows:

$$\hat{T}(s) := \sum_{i \in s^{(1)}} \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \quad (1.4)$$

is a design-unbiased estimator for

$$T = \sum_{i \in U^{(1)}} T_i$$

If the r -stage sampling design has invariant and independent subsampling, then the design-variance of \hat{T} is given by:

$$\text{Var}[\hat{T}] = \underbrace{\text{Var}^{(1)}\left(E^{(2+)}(\hat{T} \mid s^{(1)})\right)}_{V_{\text{PSU}}} + \underbrace{E^{(1)}\left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)})\right)}_{V_{\text{subsampling}}}, \quad (1.5)$$

where

$$\begin{aligned} \text{Var}^{(1)}\left(E^{(2+)}(\hat{T} \mid s^{(1)})\right) &= \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \frac{T_i}{\pi_i^{(1)}} \frac{T_j}{\pi_j^{(1)}}, \quad \text{and} \\ E^{(1)}\left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)})\right) &= \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}}, \end{aligned}$$

with

$$V_i := \text{Var}^{(2+)}[\hat{T}_i] \quad \text{and} \quad \Delta_{ij}^{(1)} := \begin{cases} \pi_i^{(1)} (1 - \pi_i^{(1)}), & \text{if } i = j \\ \pi_{ij}^{(1)} - \pi_i^{(1)} \pi_j^{(1)}, & \text{if } i \neq j \end{cases} \quad (1.6)$$

Furthermore, if $\widehat{V}_i : \mathcal{S}_i^{(2+)} \rightarrow \mathbb{R}$ is an unbiased estimator for $V_i := \text{Var}^{(2+)}[\widehat{T}_i]$, and $\pi_i^{(1)} > 0, \pi_{ij}^{(1)} > 0$ for any PSU $i, j \in U^{(1)}$, then

$$\begin{aligned} \widehat{\text{Var}}[\widehat{T}](s) &:= \sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\widehat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\widehat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} + \sum_{i \in s^{(1)}} \frac{\widehat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}}, \\ \widehat{\text{Var}}^{(1)}[\widehat{T}](s) &:= \sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\widehat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\widehat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} + \sum_{i \in s^{(1)}} \left(1 - \frac{1}{\pi_i^{(1)}}\right) \frac{\widehat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}} \\ &= \sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\widehat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\widehat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} + \sum_{i \in s^{(1)}} \frac{\widehat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}} - \sum_{i \in s^{(1)}} \frac{\widehat{V}_i(s_i^{(2+)})}{\left(\pi_i^{(1)}\right)^2} \\ \widehat{\text{Var}}^{(2+)}[\widehat{T}](s) &:= \sum_{i \in s^{(1)}} \frac{\widehat{V}_i(s_i^{(2+)})}{\left(\pi_i^{(1)}\right)^2} \end{aligned}$$

are unbiased estimators for $\text{Var}[\widehat{T}]$, $V_{\text{PSU}} := \text{Var}^{(1)}\left(E^{(2+)}(\widehat{T} \mid s^{(1)})\right)$, and $V_{\text{subsampling}} := E^{(1)}\left(\text{Var}^{(2+)}(\widehat{T} \mid s^{(1)})\right)$, respectively.

Corollary 1.2

$$\widehat{\text{Var}}^{(1)}[\widehat{T}](s) = \widehat{\text{Var}}[\widehat{T}](s) - \widehat{\text{Var}}^{(2+)}[\widehat{T}](s) \quad (1.7)$$

PROOF of Theorem 1.1

$$\begin{aligned} \text{Var}^{(1)}\left[E^{(2+)}\left[\widehat{T} \mid s^{(1)}\right]\right] &= \text{Var}^{(1)}\left[E^{(2+)}\left[\sum_{i \in s^{(1)}} \frac{\widehat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \mid s^{(1)}\right]\right] \\ &= \text{Var}^{(1)}\left[\sum_{i \in s^{(1)}} \frac{E^{(2+)}\left[\widehat{T}_i(s_i^{(2+)})\right]}{\pi_i^{(1)}}\right] \\ &= \text{Var}^{(1)}\left[\sum_{i \in s^{(1)}} \frac{T_i}{\pi_i^{(1)}}\right] \\ &= \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \frac{T_i}{\pi_i^{(1)}} \frac{T_j}{\pi_j^{(1)}} \end{aligned}$$

$$\begin{aligned}
 E^{(1)} \left[\text{Var}^{(2+)} \left[\hat{T} \mid s^{(1)} \right] \right] &= E^{(1)} \left[\text{Var}^{(2+)} \left[\sum_{i \in s^{(1)}} \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \mid s^{(1)} \right] \right] \\
 &= E^{(1)} \left[\sum_{i \in s^{(1)}} \text{Var}^{(2+)} \left[\frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \right] \right] \\
 &= E^{(1)} \left[\sum_{i \in s^{(1)}} \frac{\text{Var}^{(2+)} \left[\hat{T}_i(s_i^{(2+)}) \right]}{\left(\pi_i^{(1)} \right)^2} \right] \\
 &= E^{(1)} \left[\sum_{i \in s^{(1)}} \frac{V_i / \pi_i^{(1)}}{\pi_i^{(1)}} \right] \\
 &= \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}}
 \end{aligned}$$

$$\begin{aligned}
 E \left(\widehat{\text{Var}}^{(2+)}(\hat{T}) \right) &= E \left(\sum_{i \in s^{(1)}} \frac{\hat{V}_i(s_i^{(2+)})}{\left(\pi_i^{(1)} \right)^2} \right) = E^{(1)} \left(E^{(2+)} \left(\sum_{i \in s^{(1)}} \frac{\hat{V}_i(s_i^{(2+)}) / \pi_i^{(1)}}{\pi_i^{(1)}} \mid s^{(1)} \right) \right) \\
 &= E^{(1)} \left(\sum_{i \in s^{(1)}} \frac{E^{(2+)} \left[\hat{V}_i(s_i^{(2+)}) \mid s^{(1)} \right] / \pi_i^{(1)}}{\pi_i^{(1)}} \right) = E^{(1)} \left(\sum_{i \in s^{(1)}} \frac{V_i / \pi_i^{(1)}}{\pi_i^{(1)}} \right) = \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}} \\
 &= E^{(1)} \left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)}) \right) = V_{\text{PSU}}
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 E \left(\sum_{i \in s^{(1)}} \frac{\hat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}} \right) &= E^{(1)} \left(E^{(2+)} \left(\sum_{i \in s^{(1)}} \frac{\hat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}} \mid s^{(1)} \right) \right) = E^{(1)} \left(\sum_{i \in s^{(1)}} \frac{E^{(2+)} \left[\hat{V}_i(s_i^{(2+)}) \mid s^{(1)} \right]}{\pi_i^{(1)}} \right) \\
 &= E^{(1)} \left(\sum_{i \in s^{(1)}} \frac{V_i}{\pi_i^{(1)}} \right) = \sum_{i \in U^{(1)}} V_i
 \end{aligned}$$

Next, observe that

$$\begin{aligned}
 E \left[\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\hat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} \right] &= E^{(1)} \left(E^{(2+)} \left(\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\hat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} \mid s^{(1)} \right) \right) \\
 &= E^{(1)} \left(\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{E^{(2+)} \left[\hat{T}_i(s_i^{(2+)}) \cdot \hat{T}_j(s_j^{(2+)}) \mid s^{(1)} \right]}{\pi_i^{(1)} \pi_j^{(1)}} \right)
 \end{aligned}$$

Now, observe (the key technical observation) that

$$E^{(2+)} \left[\hat{T}_i(s_i^{(2+)}) \cdot \hat{T}_j(s_j^{(2+)}) \mid s^{(1)} \right] = E^{(2+)} \left[\hat{T}_i(s_i^{(2+)}) \cdot \hat{T}_j(s_j^{(2+)}) \right] = \begin{cases} \text{Var}^{(2+)}(\hat{T}_i) + E^{(2+)}(\hat{T}_i)^2, & \text{if } i = j, \\ E^{(2+)}(\hat{T}_i) \cdot E^{(2+)}(\hat{T}_j), & \text{if } i \neq j \end{cases}$$

Hence,

$$\begin{aligned}
 E \left[\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\hat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} \right] &= E^{(1)} \left(\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{E^{(2+)} \left[\hat{T}_i(s_i^{(2+)}) \cdot \hat{T}_j(s_j^{(2+)}) \mid s^{(1)} \right]}{\pi_i^{(1)} \pi_j^{(1)}} \right) \\
 &= E^{(1)} \left(\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{T_i}{\pi_i^{(1)}} \cdot \frac{T_j}{\pi_j^{(1)}} + \sum_{i \in s^{(1)}} \frac{(1 - \pi_i^{(1)}) V_i / \pi_i^{(1)}}{\pi_i^{(1)}} \right) \\
 &= \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \cdot \frac{T_i}{\pi_i^{(1)}} \cdot \frac{T_j}{\pi_j^{(1)}} + \sum_{i \in U^{(1)}} \frac{(1 - \pi_i^{(1)})}{\pi_i^{(1)}} \cdot V_i \\
 &= \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \cdot \frac{T_i}{\pi_i^{(1)}} \cdot \frac{T_j}{\pi_j^{(1)}} + \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}} - \sum_{i \in U^{(1)}} V_i
 \end{aligned}$$

We may now establish that

$$\begin{aligned}
 E \left[\widehat{\text{Var}}(\hat{T}) \right] &= E \left[\sum_{i \in s^{(1)}} \sum_{j \in s^{(1)}} \frac{\Delta_{ij}^{(1)}}{\pi_{ij}^{(1)}} \cdot \frac{\hat{T}_i(s_i^{(2+)})}{\pi_i^{(1)}} \cdot \frac{\hat{T}_j(s_j^{(2+)})}{\pi_j^{(1)}} + \sum_{i \in s^{(1)}} \frac{\hat{V}_i(s_i^{(2+)})}{\pi_i^{(1)}} \right] \\
 &= \left\{ \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \cdot \frac{T_i}{\pi_i^{(1)}} \cdot \frac{T_j}{\pi_j^{(1)}} + \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}} - \sum_{i \in U^{(1)}} V_i \right\} + \left\{ \sum_{i \in U^{(1)}} V_i \right\} \\
 &= \sum_{i \in U^{(1)}} \sum_{j \in U^{(1)}} \Delta_{ij}^{(1)} \cdot \frac{T_i}{\pi_i^{(1)}} \cdot \frac{T_j}{\pi_j^{(1)}} + \sum_{i \in U^{(1)}} \frac{V_i}{\pi_i^{(1)}} \\
 &= \text{Var}(\hat{T})
 \end{aligned}$$

Lastly, note that

$$\widehat{\text{Var}}^{(1)}[\hat{T}](s) = \widehat{\text{Var}}[\hat{T}](s) - \widehat{\text{Var}}^{(2+)}[\hat{T}](s)$$

Hence,

$$\begin{aligned}
 E \left[\widehat{\text{Var}}^{(1)}(\hat{T}) \right] &= E \left[\widehat{\text{Var}}(\hat{T}) \right] - E \left[\widehat{\text{Var}}^{(2+)}(\hat{T}) \right] \\
 &= \text{Var}[\hat{T}] - E^{(1)} \left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)}) \right) \\
 &= \text{Var}^{(1)} \left(E^{(2+)}(\hat{T} \mid s^{(1)}) \right) + E^{(1)} \left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)}) \right) - E^{(1)} \left(\text{Var}^{(2+)}(\hat{T} \mid s^{(1)}) \right) \\
 &= \text{Var}^{(1)} \left(E^{(2+)}(\hat{T} \mid s^{(1)}) \right) = V_{\text{subsampling}}
 \end{aligned}$$

□

2 Variance estimation for four-stage sampling, with SRSWOR at each stage

First, recall that for a simple random sampling without replacement (SRSWOR), with fixed sample size n from a population of size N , the first- and second-order selection probabilities are given by:

$$\pi_i = \frac{n}{N} \quad \text{and} \quad \pi_{ij} = \frac{n(n-1)}{N(N-1)},$$

for any distinct units i, j in the population. The Horvitz-Thompson estimator of the population total of a population characteristic y is, by definition:

$$\hat{T}_y^{\text{HT}}(s) := \frac{N}{n} \sum_{k \in s} y_k = w \cdot \sum_{k \in s} y_k, \quad \text{where } w := \frac{N}{n}.$$

The design variance of \hat{T}_y^{HT} is given by:

$$\begin{aligned} \text{Var}[\hat{T}_y^{\text{HT}}] &= \sum_{i \in U} \sum_{j \in U} \Delta_{ij} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} = \dots = N^2 \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \left\{ \frac{1}{N-1} \sum_{k \in U} (y_k - \bar{y}_U)^2 \right\} \\ &= N^2 \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \text{SVar}(\{y_k\}_{k \in U}) = (nw)^2 \left(1 - \frac{1}{w}\right) \cdot \frac{1}{n} \cdot \text{SVar}(\{y_k\}_{k \in U}) \\ &= nw(w-1) \cdot \text{SVar}(\{y_k\}_{k \in U}) = N \cdot (w-1) \cdot \text{SVar}(\{y_k\}_{k \in U}), \end{aligned}$$

where $\bar{y}_U := \frac{1}{N} \sum_{k \in U} y_k$. Recall also that a design-unbiased estimator of \hat{T}_y^{HT} is given by:

$$\begin{aligned} \widehat{\text{Var}}[\hat{T}_y^{\text{HT}}](s) &= \sum_{i \in s} \sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} = N^2 \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \left\{ \frac{1}{n-1} \sum_{k \in s} (y_k - \bar{y}_s)^2 \right\} \\ &= N^2 \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \text{SVar}(\{y_k\}_{k \in s}) \\ &= nw(w-1) \cdot \text{SVar}(\{y_k\}_{k \in s}) = N \cdot (w-1) \cdot \text{SVar}(\{y_k\}_{k \in s}) \end{aligned}$$

With the above observations, applying Theorem 1.1 recursively to $\widehat{\text{Var}}[\hat{T}]$, $\hat{V}_i^{(2+)}$, $\hat{V}_{ia}^{(3+)}$, and $\hat{V}_{iac}^{(4)}$ immediately yields the following:

Corollary 2.1

For a four-stage sampling design with invariant and independent subsampling, where sampling random sampling without replacement (SRSWOR) is used at each stage, we have

$$\begin{aligned} \widehat{\text{Var}}[\hat{T}](s) &= N^{(1)} (w^{(1)} - 1) \text{SVar}\left(\left\{\hat{T}_i\right\}_{i \in s^{(1)}}\right) + w^{(1)} \sum_{i \in s^{(1)}} \hat{V}_i^{(2+)} \\ \hat{V}_i^{(2+)} &= N_i^{(2)} (w_i^{(2)} - 1) \text{SVar}\left(\left\{\hat{T}_{ia}\right\}_{a \in s_i^{(2)}}\right) + w_i^{(2)} \sum_{a \in s_i^{(2)}} \hat{V}_{ia}^{(3+)} \\ \hat{V}_{ia}^{(3+)} &= N_{ia}^{(3)} (w_{ia}^{(3)} - 1) \text{SVar}\left(\left\{\hat{T}_{iac}\right\}_{c \in s_{ia}^{(3)}}\right) + w_{ia}^{(3)} \sum_{c \in s_{ia}^{(3)}} \hat{V}_{iac}^{(4)} \\ \hat{V}_{iac}^{(4)} &= N_{iac}^{(4)} (w_{iac}^{(4)} - 1) \text{SVar}\left(\left\{y_u\right\}_{u \in s_{iac}^{(4)}}\right) \end{aligned}$$

Multi-stage Sampling

$$\widehat{\text{Var}}^{(1)}[\widehat{T}](s) = n^{(1)}w^{(1)}(w^{(1)} - 1) \text{SVar}\left(\left\{\widehat{T}_i\right\}_{i \in s^{(1)}}\right) + w^{(1)}(1 - w^{(1)}) \sum_{i \in s^{(1)}} \widehat{V}_i^{(2+)}$$

$$\widehat{\text{Var}}^{(2+)}[\widehat{T}](s) = \left(w^{(1)}\right)^2 \sum_{i \in s^{(1)}} \widehat{V}_i^{(2+)}$$