

1 The population total, population mean, and population variance of a population characteristic

Let $n, N \in \mathbb{N}$, with $n \leq N$. Let $\mathcal{U} = \{1, 2, \dots, N\}$, which represents the finite population, or universe, of N elements.

Definition 1.1 A population characteristic is an \mathbb{R} -valued function $y : \mathcal{U} \rightarrow \mathbb{R}$ defined on the population \mathcal{U} . We denote the value of y evaluated at $i \in \mathcal{U}$ by y_i . The population total, denoted by t , of y is defined:

$$t := \sum_{i=1}^N y_i \in \mathbb{R}.$$

The population mean, denoted by \bar{y} , of y is defined by:

$$\bar{y} := \frac{1}{N} \sum_{i=1}^N y_i \in \mathbb{R}.$$

The population variance, denoted by S^2 , of y is defined by:

$$S^2 := \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 = \frac{1}{N-1} \left\{ \left(\sum_{i=1}^N y_i^2 \right) - N \cdot \bar{y}^2 \right\} \in \mathbb{R}.$$

In survey sampling, we seek to estimate population total t and population mean \bar{y} of a population characteristic $y : \mathcal{U} \rightarrow \mathbb{R}$ by making observations of values of y on only a (usually proper) subset of \mathcal{U} , and extrapolate from these observations. The subset on which observations of values of y are made is called a *sample*.

2 Simple Random Sampling (SRS)

Definition 2.1 Let \mathcal{U} be a nonempty finite set, $N := \#(\mathcal{U}) \in \mathbb{N}$, and let $n \in \{1, 2, \dots, N\}$ be given. We define the probability space $\Omega_{\text{SRS}}(\mathcal{U}, n)$ as follows: Let $\Omega(\mathcal{U}, n)$ be the set of all subsets of \mathcal{U} with n elements, i.e.

$$\Omega(\mathcal{U}, n) := \{ \omega \subset \mathcal{U} \mid \#(\omega) = n \}.$$

Note that $\#(\Omega(\mathcal{U}, n)) = \binom{N}{n}$. Let $\mathcal{P}(\Omega(\mathcal{U}, n))$ be the power set of $\Omega(\mathcal{U}, n)$. Define $\mu : \Omega \rightarrow \mathbb{R}$ to be the “uniform” probability measure on the (finite) σ -algebra $\mathcal{P}(\Omega(\mathcal{U}, n))$ determined by:

$$\mu(\omega) = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!}, \quad \text{for each } \omega \in \Omega(\mathcal{U}, n).$$

Then, $\Omega_{\text{SRS}}(\mathcal{U}, n)$ is defined to be the probability space $(\Omega(\mathcal{U}, n), \mathcal{P}(\Omega(\mathcal{U}, n)), \mu)$.

Definition 2.2 The simple-random-sampling sample total \hat{t}_{SRS} of the population characteristic y is, by definition, the random variable $\hat{t}_{\text{SRS}} : \Omega_{\text{SRS}}(\mathcal{U}, n) \rightarrow \mathbb{R}$ defined by

$$\hat{t}_{\text{SRS}}(\omega) := \frac{N}{n} \sum_{i \in \omega} y_i, \quad \text{for each } \omega \in \Omega.$$

The simple-random-sampling sample mean $\hat{\bar{y}}_{\text{SRS}}$ of the population characteristic y is, by definition, the random variable $\hat{\bar{y}}_{\text{SRS}} : \Omega_{\text{SRS}}(\mathcal{U}, n) \rightarrow \mathbb{R}$ defined by

$$\hat{\bar{y}}_{\text{SRS}}(\omega) := \frac{1}{n} \sum_{i \in \omega} y_i, \quad \text{for each } \omega \in \Omega.$$

The simple-random-sampling sample variance \hat{s}_{SRS}^2 of the population characteristic y is, by definition, the random variable $\hat{s}_{\text{SRS}}^2 : \Omega_{\text{SRS}}(\mathcal{U}, n) \rightarrow \mathbb{R}$ defined by

$$\hat{s}_{\text{SRS}}^2(\omega) := \frac{1}{n-1} \sum_{i \in \omega} \left(y_i - \hat{\bar{y}}_{\text{SRS}}(\omega) \right)^2, \quad \text{for each } \omega \in \Omega.$$

Proposition 2.3

1. $\hat{\bar{y}}_{\text{SRS}}$ is an unbiased estimator of the population mean \bar{y} , and $\text{Var}[\hat{\bar{y}}_{\text{SRS}}] = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$.
2. \hat{t}_{SRS} is an unbiased estimator of the population total t , and $\text{Var}[\hat{t}_{\text{SRS}}] = N^2 \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$.
3. \hat{s}_{SRS}^2 is an unbiased estimator of the population variance S^2 .
4. $\widehat{\text{Var}}[\hat{\bar{y}}_{\text{SRS}}] := \left(1 - \frac{n}{N}\right) \frac{\hat{s}_{\text{SRS}}^2}{n}$ is an unbiased estimator of $\text{Var}[\hat{\bar{y}}_{\text{SRS}}]$.
5. $\widehat{\text{Var}}[\hat{t}_{\text{SRS}}] := N^2 \left(1 - \frac{n}{N}\right) \frac{\hat{s}_{\text{SRS}}^2}{n}$ is an unbiased estimator of $\text{Var}[\hat{t}_{\text{SRS}}]$.

A quote from Lohr [2], p.37: *Hájek [1] proves a central limit theorem for simple random sampling without replacement. In practical terms, Hájek's theorem says that if certain technical conditions hold, and if n , N , and $N - n$ are all "sufficiently large," then the sampling distribution of*

$$\frac{\hat{\bar{y}}_{\text{SRS}} - \bar{y}}{\sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}}}$$

is "approximately" normal (Gaussian) with mean 0 and variance 1.

Corollary 2.4 (to Hájek's theorem) For a simple random sampling procedure, an approximate $(1 - \alpha)$ -confidence interval, $0 < \alpha < 1$, for the population mean \bar{y} is given by:

$$\hat{\bar{y}}_{\text{SRS}} \pm z_{\alpha/2} \cdot \sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}}$$

For sufficiently large samples, the above approximate confidence interval can itself be estimated from observations by:

$$\hat{\bar{y}}_{\text{SRS}} \pm \text{SE}[\hat{\bar{y}}_{\text{SRS}}] = \hat{\bar{y}}_{\text{SRS}} \pm \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{s}_{\text{SRS}}^2}{n}}$$

where

$$\text{SE}[\hat{\bar{y}}_{\text{SRS}}] := \sqrt{\widehat{\text{Var}}[\hat{\bar{y}}_{\text{SRS}}]} = \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{s}_{\text{SRS}}^2}{n}}$$

In order to prove Proposition 2.3, we introduce some auxiliary random variables:

Definition 2.5 Let $n, N \in \mathbb{N}$, with $n < N$, $\mathcal{U} := \{1, 2, \dots, N\}$, and $\Omega := \{\omega \subset \mathcal{U} \mid \#(\omega) = n\}$. For each $i \in \mathcal{U} = \{1, 2, \dots, N\}$, we define the random variable $Z_i : \Omega \rightarrow \{0, 1\}$ as follows:

$$Z_i(\omega) = \begin{cases} 1, & \text{if } i \in \omega, \\ 0, & \text{if } i \notin \omega \end{cases}.$$

Immediate observations:

- $\hat{t}_{\text{SRS}} = \frac{N}{n} \sum_{i=1}^N Z_i y_i$, as random variables on (Ω, P) , i.e.

$$\hat{t}_{\text{SRS}}(\omega) = \frac{N}{n} \sum_{i=1}^N Z_i(\omega) y_i, \quad \text{for each } \omega \in \Omega.$$

- $\hat{y}_{\text{SRS}} = \frac{1}{n} \sum_{i=1}^N Z_i y_i$, as random variables on (Ω, P) , i.e.

$$\hat{y}_{\text{SRS}}(\omega) = \frac{1}{n} \sum_{i=1}^N Z_i(\omega) y_i, \quad \text{for each } \omega \in \Omega.$$

- $E[Z_i] = \frac{n}{N}$. Indeed,

$$E[Z_i] = 1 \cdot P(Z_i = 1) + 0 \cdot P(Z_i = 0) = P(Z_i = 1) = \frac{\text{number of samples containing } i}{\text{number of all possible samples}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

- $Z_i^2 = Z_i$, since $\text{range}(Z_i) = \{0, 1\}$. Consequently,

$$E[Z_i^2] = E[Z_i] = \frac{n}{N}.$$

- $\text{Var}[Z_i] = \frac{n}{N} \left(1 - \frac{n}{N}\right)$. Indeed,

$$\begin{aligned} \text{Var}[Z_i] &:= E[(Z_i - E[Z_i])^2] = E[Z_i^2] - (E[Z_i])^2 \\ &= E[Z_i] - \left(\frac{n}{N}\right)^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 \\ &= \frac{n}{N} \left(1 - \frac{n}{N}\right). \end{aligned}$$

- For $i \neq j$, we have $E[Z_i \cdot Z_j] = \left(\frac{n-1}{N-1}\right) \cdot \left(\frac{n}{N}\right)$. Indeed,

$$\begin{aligned} E[Z_i \cdot Z_j] &= 1 \cdot P(Z_i = 1 \text{ and } Z_j = 1) + 0 \cdot P(Z_i = 0 \text{ or } Z_j = 0) \\ &= P(Z_i = 1 \text{ and } Z_j = 1) = P(Z_j = 1 | Z_i = 1) \cdot P(Z_i = 1) \\ &= \left(\frac{n-1}{N-1}\right) \cdot \left(\frac{n}{N}\right) \end{aligned}$$

- For $i \neq j$, we have $\text{Cov}(Z_i, Z_j) = -\frac{1}{N-1} \left(1 - \frac{n}{N}\right) \left(\frac{n}{N}\right) \leq 0$. Indeed,

$$\begin{aligned} \text{Cov}(Z_i, Z_j) &:= E[(Z_i - E[Z_i]) \cdot (Z_j - E[Z_j])] = E[Z_i Z_j] - E[Z_i] \cdot E[Z_j] \\ &= \left(\frac{n-1}{N-1}\right) \cdot \left(\frac{n}{N}\right) - \left(\frac{n}{N}\right)^2 = \frac{n}{N} \left(\frac{nN - N - nN + n}{N(N-1)}\right) \\ &= -\frac{1}{N-1} \left(1 - \frac{n}{N}\right) \left(\frac{n}{N}\right) \end{aligned}$$

PROOF OF Proposition 2.3

1.

$$E\left[\widehat{\bar{y}}_{\text{SRS}}\right] = E\left[\frac{1}{n} \sum_{i=1}^N Z_i y_i\right] = \frac{1}{n} \sum_{i=1}^N E[Z_i] \cdot y_i = \frac{1}{n} \sum_{i=1}^N \left(\frac{n}{N}\right) \cdot y_i = \frac{1}{N} \sum_{i=1}^N y_i =: \bar{y}.$$

$$\begin{aligned} \text{Var}\left[\widehat{\bar{y}}_{\text{SRS}}\right] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^N Z_i y_i\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^N Z_i y_i\right] = \frac{1}{n^2} \text{Cov}\left[\sum_{i=1}^N Z_i y_i, \sum_{j=1}^N Z_j y_j\right] \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^N y_i^2 \text{Var}(Z_i) + \sum_{i=1}^N \sum_{i \neq j=1}^N y_i y_j \text{Cov}(Z_i, Z_j) \right\} \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^N y_i^2 \frac{n}{N} \left(1 - \frac{n}{N}\right) - \sum_{i=1}^N \sum_{i \neq j=1}^N y_i y_j \frac{1}{N-1} \left(1 - \frac{n}{N}\right) \left(\frac{n}{N}\right) \right\} \\ &= \frac{1}{n^2} \frac{n}{N} \left(1 - \frac{n}{N}\right) \left\{ \sum_{i=1}^N y_i^2 - \frac{1}{N-1} \sum_{i=1}^N \sum_{i \neq j=1}^N y_i y_j \right\} \\ &= \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N(N-1)} \left\{ (N-1) \sum_{i=1}^N y_i^2 - \sum_{i=1}^N \sum_{i \neq j=1}^N y_i y_j \right\} \\ &= \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N(N-1)} \left\{ (N-1) \sum_{i=1}^N y_i^2 - \sum_{i=1}^N \sum_{j=1}^N y_i y_j + \sum_{i=1}^N y_i^2 \right\} \\ &= \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N(N-1)} \left\{ N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i\right) \left(\sum_{j=1}^N y_j\right) \right\} \\ &= \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N-1} \left\{ \sum_{i=1}^N y_i^2 - N \left(\frac{1}{N} \sum_{i=1}^N y_i\right)^2 \right\} \\ &= \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N-1} \left\{ \sum_{i=1}^N y_i^2 - N \cdot \bar{y}^2 \right\} \\ &= \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \end{aligned}$$

2.

$$\begin{aligned} E\left[\widehat{t}_{\text{SRS}}\right] &= E\left[N \cdot \widehat{\bar{y}}_{\text{SRS}}\right] = N \cdot E\left[\widehat{\bar{y}}_{\text{SRS}}\right] = N \cdot \bar{y} = N \cdot \left(\frac{1}{N} \sum_{i=1}^N y_i\right) = \sum_{i=1}^N y_i =: t. \\ \text{Var}\left[\widehat{t}_{\text{SRS}}\right] &= \text{Var}\left[N \cdot \widehat{\bar{y}}_{\text{SRS}}\right] = N^2 \cdot \text{Var}\left[\widehat{\bar{y}}_{\text{SRS}}\right] = N^2 \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \end{aligned}$$

3.

$$\begin{aligned}
 E\left[\widehat{s}_{\text{SRS}}^2\right] &= E\left[\frac{1}{n-1} \sum_{i \in \omega} (y_i - \widehat{\bar{y}}_{\text{SRS}})^2\right] = \frac{1}{n-1} E\left[\sum_{i \in \omega} \left((y_i - \bar{y}) - (\widehat{\bar{y}}_{\text{SRS}} - \bar{y})\right)^2\right] \\
 &= \frac{1}{n-1} E\left[\left(\sum_{i \in \omega} (y_i - \bar{y})^2\right) - n(\widehat{\bar{y}}_{\text{SRS}} - \bar{y})^2\right] \\
 &= \frac{1}{n-1} \left\{ E\left[\sum_{i=1}^N Z_i (y_i - \bar{y})^2\right] - n \text{Var}\left[\widehat{\bar{y}}_{\text{SRS}}\right] \right\} \\
 &= \frac{1}{n-1} \left\{ \sum_{i=1}^N E[Z_i] (y_i - \bar{y})^2 - n \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \right\} \\
 &= \frac{1}{n-1} \left\{ \sum_{i=1}^N \frac{n}{N} (y_i - \bar{y})^2 - \left(1 - \frac{n}{N}\right) S^2 \right\} \\
 &= \frac{1}{n-1} \left\{ \frac{n(N-1)}{N} \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 - \left(1 - \frac{n}{N}\right) S^2 \right\} \\
 &= \frac{1}{n-1} \left\{ \frac{n(N-1)}{N} - \left(1 - \frac{n}{N}\right) \right\} S^2 \\
 &= \frac{1}{n-1} \left\{ \frac{nN - n - N + n}{N} \right\} S^2 = S^2
 \end{aligned}$$

4. Immediate from preceding statements.

5. Immediate from preceding statements. □

3 Stratified Simple Random Sampling

Let $\mathcal{U} = \{1, 2, \dots, N\}$ be the population, as before. Let

$$\mathcal{U} = \bigsqcup_{h=1}^H \mathcal{U}_h$$

be a partition of \mathcal{U} . Such a partition is called a *stratification* of the population \mathcal{U} . Each of $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_H$ is called a *stratum*. Let $N_h := \#(\mathcal{U}_h)$, for $h = 1, 2, \dots, H$. Note that $N_1 + N_2 + \dots + N_H = N$.

In *stratified simple random sampling*, an SRS is taken within each stratum \mathcal{U}_h , $h = 1, 2, \dots, H$. Let n_h , $h = 1, 2, \dots, H$, be the number elements in the simple random sample taken in the stratum \mathcal{U}_h . In other words, a stratified simple random sample ω of the stratified population $\mathcal{U} = \bigsqcup_{h=1}^H \mathcal{U}_h$ has the form:

$$\omega = \bigsqcup_{h=1}^H \omega_h, \quad \text{where } \omega_h \in \Omega_{\text{SRS}}(\mathcal{U}_h, n_h), \quad \text{for each } h = 1, 2, \dots, H.$$

Note that $n_1 + n_2 + \dots + n_H =: n = \#(\omega)$.

We now give unbiased estimators, and their variances, of the population total and population mean of a population characteristic under stratified simple random sampling. Let $y : \mathcal{U} \rightarrow \mathbb{R}$ be a population characteristic. Define:

$$\begin{aligned}
 \widehat{t}_{\text{Str}} &:= \sum_{h=1}^H N_h \cdot \widehat{\bar{y}}_{h, \text{SRS}} \\
 \widehat{\bar{y}}_{\text{Str}} &:= \frac{1}{N} \cdot \widehat{t}_{\text{Str}} = \sum_{h=1}^H \frac{N_h}{N} \cdot \widehat{\bar{y}}_{h, \text{SRS}}
 \end{aligned}$$

Here,

$$\widehat{y}_{h,\text{SRS}} : \Omega_{\text{SRS}}(\mathcal{U}_h, n_h) \longrightarrow \mathbb{R} : \omega_h \longmapsto \frac{1}{n_h} \sum_{i \in \omega_h} y_i$$

is the SRS estimator of

$$\bar{y}_h := \overline{y|_{\mathcal{U}_h}} = \frac{1}{N_h} \sum_{i \in \mathcal{U}_h} y_i \in \mathbb{R},$$

the “stratum mean” of the “stratum characteristic” $y|_{\mathcal{U}_h} : \mathcal{U}_h \longrightarrow \mathbb{R}$, the restriction of the population characteristic $y : \mathcal{U} \longrightarrow \mathbb{R}$ to the stratum \mathcal{U}_h . Then,

$$E[\widehat{t}_{\text{Str}}] = t := \sum_{i=1}^N y_i, \quad \text{and} \quad E[\widehat{\bar{y}}_{\text{Str}}] = \bar{y} := \frac{1}{N} \sum_{i=1}^N y_i.$$

In other words, \widehat{t}_{Str} and $\widehat{\bar{y}}_{\text{Str}}$ are unbiased estimators of the population total t and population mean \bar{y} of the population characteristic $y : \mathcal{U} \longrightarrow \mathbb{R}$, respectively. Indeed,

$$\begin{aligned} E[\widehat{t}_{\text{Str}}] &= E\left[\sum_{h=1}^H N_h \cdot \widehat{y}_{h,\text{SRS}}\right] = \sum_{h=1}^H N_h E[\widehat{y}_{h,\text{SRS}}] = \sum_{h=1}^H N_h \bar{y}_h \\ &= \sum_{h=1}^H N_h \left(\frac{1}{N_h} \sum_{i \in \mathcal{U}_h} y_i\right) = \sum_{h=1}^H \left(\sum_{i \in \mathcal{U}_h} y_i\right) = \sum_{i=1}^N y_i =: t. \end{aligned}$$

And,

$$E[\widehat{\bar{y}}_{\text{Str}}] = E\left[\frac{1}{N} \cdot \widehat{t}_{\text{Str}}\right] = \frac{1}{N} E[\widehat{t}_{\text{Str}}] = \frac{1}{N} \sum_{i=1}^N y_i =: \bar{y}.$$

Furthermore,

$$\text{Var}[\widehat{t}_{\text{Str}}] = \text{Var}\left[\sum_{h=1}^H N_h \cdot \widehat{y}_{h,\text{SRS}}\right] = \sum_{h=1}^H N_h^2 \cdot \text{Var}[\widehat{y}_{h,\text{SRS}}] = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}.$$

$$\text{Var}[\widehat{\bar{y}}_{\text{Str}}] = \text{Var}\left[\frac{1}{N} \cdot \widehat{t}_{\text{Str}}\right] = \frac{1}{N^2} \cdot \text{Var}[\widehat{t}_{\text{Str}}] = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}.$$

Comparing variances of SRS and stratified simple random sampling with proportional allocation via ANOVA (analysis of variance):

By definition, in stratified simple random sampling with proportional allocation, the stratum sample size n_h , for each $h = 1, 2, \dots, H$, is chosen such that $n_h/N_h = n/N$. Consequently,

$$\begin{aligned} \text{Var}[\widehat{t}_{\text{PropStr}}] &= \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h} = \frac{N}{n} \left(1 - \frac{n}{N}\right) \sum_{h=1}^H N_h S_h^2 \\ &= \frac{N}{n} \left(1 - \frac{n}{N}\right) \left\{ \sum_{h=1}^H (N_h - 1) S_h^2 + \sum_{h=1}^H S_h^2 \right\} \\ &= \frac{N}{n} \left(1 - \frac{n}{N}\right) \left\{ \text{SSW} + \sum_{h=1}^H S_h^2 \right\}, \end{aligned}$$

where

$$\text{SSW} := \sum_{h=1}^H \sum_{i \in \mathcal{U}_h} (y_i - \bar{y}_{\mathcal{U}_h})^2 = \sum_{h=1}^H (N_h - 1) S_h^2.$$

is called the *inter-strata squared deviation* (or *within-strata squared deviation*), and

$$S_h^2 := \frac{1}{N_h - 1} \sum_{i \in \mathcal{U}_h} (y_i - \bar{y}_{\mathcal{U}_h})^2$$

is the stratum variance of the population characteristic $y : \mathcal{U} \rightarrow \mathbb{R}$ over the stratum \mathcal{U}_h . The following relation between $\text{Var}[\hat{t}_{\text{SRS}}]$ and $\text{Var}[\hat{t}_{\text{PropStr}}]$ always holds (see [2], p.106):

$$\text{Var}[\hat{t}_{\text{SRS}}] = \text{Var}[\hat{t}_{\text{PropStr}}] + \left(1 - \frac{n}{N}\right) \frac{N}{n} \frac{N}{N-1} \left\{ \text{SSB} - \sum_{h=1}^H \left(1 - \frac{N_h}{N}\right) S_h^2 \right\},$$

where

$$\text{SSB} := \sum_{h=1}^H N_h (\bar{y}_{\mathcal{U}_h} - \bar{y}_{\mathcal{U}})^2 = \sum_{h=1}^H \sum_{i \in \mathcal{U}_h} (\bar{y}_{\mathcal{U}_h} - \bar{y}_{\mathcal{U}})^2$$

is the *inter-strata squared deviation* (or *between-strata squared deviation*). It is also an easily established fact that the sum of the inter-strata squared deviation SSB and the intra-strata squared deviation SSW is always the total population squared deviation SSTO:

$$\text{SSTO} := \sum_{i=1}^N (y_i - \bar{y}_{\mathcal{U}})^2 = \sum_{h=1}^H \sum_{i \in \mathcal{U}_h} (y_i - \bar{y}_{\mathcal{U}})^2.$$

Most importantly, we see from above that, for stratified simple random sampling with proportional allocation, the following implication holds:

$$\sum_{h=1}^H \left(1 - \frac{N_h}{N}\right) S_h^2 \leq \text{SSB} \implies \text{Var}[\hat{t}_{\text{PropStr}}] \leq \text{Var}[\hat{t}_{\text{SRS}}].$$

In heuristic terms, in proportional-allocation stratification for which each stratum is relatively homogeneous and the strata are relatively dissimilar to each other (intra-strata variation being smaller than inter-strata variation), then the unbiased estimator for the population total from the proportional-allocation stratified simple random sampling is more precise than that from SRS.

4 Two-stage Cluster Sampling

The universe $\mathcal{U} = \bigsqcup_{i=1}^N \mathcal{C}_i$ of observation units is partitioned into N clusters (or *primary sampling units*, psu's) \mathcal{C}_i . In two-stage cluster sampling, the *secondary sampling units* (or ssu's) are the observation units. Let M_i be the number of ssu's in the i th psu; in other words, $M_i := \#(\mathcal{C}_i)$.

First Stage: Select a simple random sample (SRS) $\omega_1 = \{\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \dots, \mathcal{C}_{i_n}\}$ of n psu's from the collection of N psu's.

Second Stage: From each psu $\mathcal{C} \in \omega_1$ selected in the First Stage, select a simple random sample (SRS) $\omega_{\mathcal{C}}$ of m_i secondary sampling units (ssu's) from the collection of M_i ssu's in \mathcal{C} .

The sample is then $\omega := \bigsqcup_{\mathcal{C} \in \omega_1} \omega_{\mathcal{C}}$. In other words, the sample ω consists of all the secondary sampling units (or observation units) selected (during the Second Stage) from all the primary sampling units selected in the First Stage.

The Horvitz-Thompson estimator \hat{t}_{HT} , as defined below, is an unbiased estimator for the total of an \mathbb{R} -valued population characteristic $y : \mathcal{U} \rightarrow \mathbb{R}$.

$$\hat{t}_{\text{HT}} := \sum_{k \in \omega} \left(\frac{N}{n} \frac{M_{y_k}}{m_{y_k}} \right) y_k = \sum_{k \in \omega} \left(\frac{1}{\pi_k} \right) y_k = \sum_{\mathcal{C} \in \omega_1} \sum_{k \in \omega_{\mathcal{C}}} \left(\frac{N}{n} \frac{M_{y_k}}{m_{y_k}} \right) y_k,$$

where $M_{y_k} := M_i := \#(\mathcal{C}_i)$ and $m_{y_k} := m_i := \#(\omega_{\mathcal{C}_i})$ such that \mathcal{C}_i is the unique psu containing the ssu $k \in \mathcal{U} = \bigsqcup_i^N \mathcal{C}_i$.

The variance of the Horvitz-Thompson estimator \hat{t}_{HT} is given by:

$$\text{Var}(\hat{t}_{\text{HT}}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} + \sum_{i=1}^N \frac{N}{n} \cdot M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{S_i^2}{m_i},$$

where

$$S_t^2 := \frac{1}{N-1} \sum_{i=1}^N \left(t_i - \frac{t}{N}\right)^2, \quad S_i^2 := \frac{1}{M_i-1} \sum_{j=1}^{M_i} \left(y_j - \frac{t_i}{M_i}\right)^2, \quad t := \sum_{k \in \mathcal{U}} y_k, \quad \text{and} \quad t_i := \sum_{k \in \mathcal{C}_i} y_k$$

IMPORTANT OBSERVATION: The first summand in the expression of $\text{Var}(\hat{t}_{\text{HT}})$ is due to variability in the First-Stage sampling, whereas the second summand is due to variability in the Second-Stage sampling.

5 One-stage Cluster Sampling

One-stage cluster sampling is a special form of two-stage cluster sampling in which all second-stage samples are censuses. In other words, following the notation introduced for two-stage cluster sampling, in one-stage cluster sampling, we have $\omega_{\mathcal{C}} = \mathcal{C}$, for each first-stage-selected $\mathcal{C} \in \omega_1$. This also implies $m_i = M_i$ for each $i = 1, 2, \dots, N$.

Then, the Horvitz-Thompson estimator \hat{t}_{HT} and its variance reduces to:

$$\begin{aligned} \hat{t}_{\text{HT}} &:= \sum_{\mathcal{C} \in \omega_1} \sum_{k \in \mathcal{C}} \left(\frac{N}{n} \frac{M_{y_k}}{m_{y_k}}\right) y_k = \frac{N}{n} \cdot \sum_{\mathcal{C} \in \omega_1} \sum_{k \in \mathcal{C}} y_k = \frac{N}{n} \cdot \sum_{\mathcal{C} \in \omega_1} t_{\mathcal{C}}, \quad \text{where } t_{\mathcal{C}} := \sum_{k \in \mathcal{C}} y_k \\ \text{Var}(\hat{t}_{\text{HT}}) &= N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} + \sum_{i=1}^N \frac{N}{n} \cdot M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{S_i^2}{m_i} \\ &= N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} + \sum_{i=1}^N \frac{N}{n} \cdot M_i^2 (1 - 1) \frac{S_i^2}{m_i} = N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} \end{aligned}$$

6 Stratified Simple Random Sampling as a special case of Two-stage Cluster Sampling

Stratified simple random sampling is a special case of two-stage cluster sampling in which the first-stage sampling is a census. In other words, if $\mathcal{U} = \bigsqcup_{i=1}^N \mathcal{C}_i$, then $\omega_1 = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_N\}$. In particular, $n = N$.

Then, the Horvitz-Thompson estimator \hat{t}_{HT} and its variance reduces to:

$$\begin{aligned} \hat{t}_{\text{HT}} &:= \sum_{\mathcal{C} \in \omega_1} \sum_{k \in \omega_{\mathcal{C}}} \left(\frac{N}{n} \frac{M_{y_k}}{m_{y_k}}\right) y_k = \sum_{i=1}^N M_i \left(\frac{1}{m_i} \sum_{k \in \omega_{\mathcal{C}_i}} y_k\right) = \sum_{i=1}^N M_i \bar{y}_{\omega_{\mathcal{C}_i}} \\ \text{Var}(\hat{t}_{\text{HT}}) &= N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} + \sum_{i=1}^N \frac{N}{n} \cdot M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{S_i^2}{m_i} \\ &= N^2 (1 - 1) \frac{S_t^2}{n} + \sum_{i=1}^N 1 \cdot M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{S_i^2}{m_i} = \sum_{i=1}^N M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{S_i^2}{m_i} \end{aligned}$$

The above formula agree exactly with those derived earlier for stratified simple random sampling (apart from obvious notational changes).

7 General linear estimators for (multivariate) population totals

Let $U = \{1, 2, \dots, N\}$ be a finite population. Let $\mathbf{y} : U \rightarrow \mathbb{R}^m$ be an \mathbb{R}^m -valued function defined on U (commonly called a “population parameter”). We will use the common notation \mathbf{y}_k for $\mathbf{y}(k)$. We wish to estimate $\mathbf{T}_{\mathbf{y}} := \sum_{k \in U} \mathbf{y}_k \in \mathbb{R}^m$ via survey sampling. Let $p : \mathcal{S} \rightarrow (0, 1]$ be our chosen sampling design, where $\mathcal{S} \subseteq \mathcal{P}(U)$ is the set of all possible samples in the design, and $\mathcal{P}(U)$ is the power set of U .

Definition 7.1

A random variable $\hat{\mathbf{T}}_{\mathbf{y}} : \mathcal{S} \rightarrow \mathbb{R}^m$ is said to be linear in the population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$ if it has the following form:

$$\begin{aligned} \hat{\mathbf{T}}_{\mathbf{y}} : \mathcal{S} &\rightarrow \mathbb{R}^m \\ s &\mapsto \sum_{k \in s} w_k(s) \mathbf{y}_k = \sum_{k \in U} I_k(s) w_k(s) \mathbf{y}_k, \end{aligned}$$

where, for each $k \in U$, $w_k : \mathcal{S} \rightarrow \mathbb{R}$ is itself an \mathbb{R} -valued random variable, and $I_k : \mathcal{S} \rightarrow \{0, 1\}$ is the indicator random variable defined by:

$$I_k(s) = \begin{cases} 1, & \text{if } k \in s, \\ 0, & \text{otherwise} \end{cases}$$

We call the w_k 's the weights of $\hat{\mathbf{T}}_{\mathbf{y}}$, and we use the notation $\hat{\mathbf{T}}_{\mathbf{y};w}$ to indicate that the random variable depends on the weights w_k .

Nomenclature In the context of finite-population probability sampling, under a design $p : \mathcal{S} \rightarrow (0, 1]$, an “estimator” is precisely just a random variable defined on the space \mathcal{S} of all admissible samples in the design.

Proposition 7.2

Let $\hat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S} \rightarrow \mathbb{R}^m$, with $\hat{\mathbf{T}}_{\mathbf{y};w}(s) = \sum_{k \in s} I_k(s) w_k(s) \mathbf{y}_k = \sum_{k \in U} w_k(s) \mathbf{y}_k$, be a random variable linear in the population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$. Then,

$$E[\hat{\mathbf{T}}_{\mathbf{y};w}] = \mathbf{T}_{\mathbf{y}}, \text{ for arbitrary } \mathbf{y} \iff E[I_k w_k] = 1, \text{ for each } k \in U.$$

PROOF Note:

$$E[\hat{\mathbf{T}}_{\mathbf{y};w}] = E\left[\sum_{k \in s} w_k \mathbf{y}_k\right] = E\left[\sum_{k \in U} I_k w_k \mathbf{y}_k\right] = \sum_{k \in U} E[I_k w_k] \mathbf{y}_k$$

Hence, since $\mathbf{y} : U \rightarrow \mathbb{R}^m$ is arbitrary,

$$E[\hat{\mathbf{T}}_{\mathbf{y};w}] = \mathbf{T}_{\mathbf{y}} := \sum_{k \in U} \mathbf{y}_k \iff \sum_{k \in U} (E[I_k w_k] - 1) \cdot \mathbf{y}_k = \mathbf{0} \iff E[I_k w_k] = 1, \text{ for each } k \in U.$$

The proof of the Proposition is now complete. □

Corollary 7.3

Let $U = \{1, 2, \dots, N\}$ be a finite population. For any fixed but arbitrary population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$ and for any sampling design $p : \mathcal{S} \rightarrow (0, 1]$ such that each of its first-order inclusion probabilities is strictly positive, the Horvitz-Thompson estimator $\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}$ is well-defined and it is the unique unbiased estimator for $\mathbf{T}_{\mathbf{y}}$, which is linear in \mathbf{y} and whose weights are constant in s .

PROOF Recall that the Horvitz-Thompson estimator is defined as:

$$\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}(s) := \sum_{k \in s} \frac{1}{\pi_k} \mathbf{y}_k := \sum_{k \in U} I_k(s) \frac{1}{\pi_k} \mathbf{y}_k,$$

where $\pi_k := E[I_k] = \sum_{k \in U} p(s) I_k(s) = \sum_{s \ni k} p(s)$ is the inclusion probability of $k \in U$ under the sampling design $p : \mathcal{S} \rightarrow (0, 1]$. Clearly, $\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}$ is linear in \mathbf{y} with weights constant in s . Next, note that:

$$E[\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}] = E\left[\sum_{k \in s} \frac{1}{\pi_k} \mathbf{y}_k\right] = E\left[\sum_{k \in U} I_k \frac{\mathbf{y}_k}{\pi_k}\right] = \sum_{k \in U} E[I_k] \frac{\mathbf{y}_k}{\pi_k} = \sum_{k \in U} \pi_k \frac{\mathbf{y}_k}{\pi_k} = \sum_{k \in U} \mathbf{y}_k = \mathbf{T}_y$$

Hence, $\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}$ is an unbiased estimator for \mathbf{T}_y . Conversely, let

$$\hat{\mathbf{T}}_{y;w}(s) = \sum_{k \in s} w_k \mathbf{y}_k$$

be any unbiased estimator for \mathbf{T}_y which linear in \mathbf{y} with weights w_k constant in s . Thus,

$$\sum_{k \in U} \mathbf{y}_k = \mathbf{T}_y = E[\hat{\mathbf{T}}_{y;w}] = E\left[\sum_{k \in s} w_k \mathbf{y}_k\right] = E\left[\sum_{k \in U} I_k w_k \mathbf{y}_k\right] = \sum_{k \in U} E[I_k] w_k \mathbf{y}_k = \sum_{k \in U} \pi_k w_k \mathbf{y}_k.$$

Since \mathbf{y} is arbitrary, the above equation immediately implies that

$$\pi_k w_k - 1 = 0,$$

or equivalently, $w_k = \frac{1}{\pi_k}$; in other words, $\hat{\mathbf{T}}_{y;w}$ is in fact equal to the Horvitz-Thompson estimator. The proof of the Corollary is now complete. \square

Lemma 7.4

Let (Ω, \mathcal{A}, p) be a probability space, $X, Y : \Omega \rightarrow \mathbb{R}$ be two \mathbb{R} -valued random variables defined on Ω , and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ be two fixed vectors in \mathbb{R}^m . Then,

$$\text{Cov}(X \cdot \mathbf{u}, Y \cdot \mathbf{v}) = \text{Cov}(X, Y) \cdot \mathbf{u} \cdot \mathbf{v}^T \in \mathbb{R}^{m \times m}$$

PROOF Note:

$$\begin{aligned} \text{Cov}(X \cdot \mathbf{u}, Y \cdot \mathbf{v}) &:= E[(X \mathbf{u} - \mu_X \mathbf{u}) \cdot (Y \mathbf{v} - \mu_Y \mathbf{v})^T] = E[(X - \mu_X) \mathbf{u} \cdot (Y - \mu_Y) \mathbf{v}^T] \\ &= E[(X - \mu_X)(Y - \mu_Y) \cdot \mathbf{u} \cdot \mathbf{v}^T] = E[(X - \mu_X)(Y - \mu_Y)] \cdot \mathbf{u} \cdot \mathbf{v}^T \\ &= \text{Cov}(X, Y) \cdot \mathbf{u} \cdot \mathbf{v}^T, \end{aligned}$$

as required. \square

Proposition 7.5

Let $\hat{\mathbf{T}}_{y;w} : \mathcal{S} \rightarrow \mathbb{R}$, with $\hat{\mathbf{T}}_{y;w}(s) = \sum_{k \in s} w_k(s) \mathbf{y}_k = \sum_{k \in U} I_k(s) w_k(s) \mathbf{y}_k$, be a random variable linear in the population parameter $\mathbf{y} : U \rightarrow \mathbb{R}$. Then, the covariance matrix of $\hat{\mathbf{T}}_{y;w}$ is given by:

$$\text{Var}[\hat{\mathbf{T}}_{y;w}] = \sum_{i \in U} \sum_{k \in U} \text{Cov}[I_i w_i, I_k w_k] \mathbf{y}_i \cdot \mathbf{y}_k^T \in \mathbb{R}^{m \times m}$$

Furthermore, if the first-order and second-order inclusion probabilities of the sampling design $p : \mathcal{S} \rightarrow (0, 1]$ are all strictly positive, i.e. $\pi_k = \pi_{kk} := \sum_{s \ni k} p(s) > 0$, for each $k \in U$, and $\pi_{ik} := \sum_{s \ni i, k} p(s) > 0$, for any distinct $i, k \in U$, then

an unbiased estimator for $\text{Var}[\hat{\mathbf{T}}_{y;w}]$ is given by:

$$\widehat{\text{Var}}[\hat{\mathbf{T}}_{y;w}](s) := \sum_{i, k \in s} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \mathbf{y}_i \cdot \mathbf{y}_k^T = \sum_{k \in s} \frac{\text{Var}(I_k w_k)}{\pi_k} \mathbf{y}_k \cdot \mathbf{y}_k^T + \sum_{\substack{i, k \in s \\ i \neq k}} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \mathbf{y}_i \cdot \mathbf{y}_k^T, \text{ for each } s \in \mathcal{S}.$$

PROOF First, note that Lemma 7.4 implies:

$$\text{Var}[\widehat{\mathbf{T}}_{\mathbf{y};w}] = \text{Cov}\left[\sum_{i \in U} I_i w_i \mathbf{y}_i, \sum_{k \in U} I_k w_k \mathbf{y}_k\right] = \sum_{i \in U} \sum_{k \in U} \text{Cov}[I_i w_i, I_k w_k] \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \in \mathbb{R}^{m \times m}$$

Next,

$$\begin{aligned} E\left(\widehat{\text{Var}}[\widehat{\mathbf{T}}_{\mathbf{y};w}]\right) &= \sum_{s \in \mathcal{S}} p(s) \cdot \widehat{\text{Var}}[\widehat{\mathbf{T}}_{\mathbf{y};w}](s) = \sum_{s \in \mathcal{S}} p(s) \cdot \left(\sum_{i,k \in s} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \right) \\ &= \sum_{s \in \mathcal{S}} p(s) \cdot \left(\sum_{i,k \in U} I_i(s) I_k(s) \cdot \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \right) \\ &= \sum_{i,k \in U} \left(\sum_{s \in \mathcal{S}} p(s) I_i(s) I_k(s) \right) \cdot \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\ &= \sum_{i,k \in U} \left(\sum_{s \ni i,k} p(s) \right) \cdot \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\ &= \sum_{i,k \in U} \pi_{ik} \cdot \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T = \sum_{i,k \in U} \text{Cov}(I_i w_i, I_k w_k) \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\ &= \text{Var}[\widehat{\mathbf{T}}_{\mathbf{y};w}] \end{aligned}$$

Lastly, recall that $\pi_{kk} = \pi_k$ and $\text{Cov}(I_k w_k, I_k w_k) = \text{Var}[I_k w_k]$, and the validity of the following identity is thus trivial:

$$\sum_{i,k \in s} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T = \sum_{k \in s} \frac{\text{Var}(I_k w_k)}{\pi_k} \cdot \mathbf{y}_k \cdot \mathbf{y}_k^T + \sum_{\substack{i,k \in s \\ i \neq k}} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T$$

The proof of the Proposition is complete. □

8 Unbiased variance estimators for the Horvitz-Thompson Estimator

Let $U = \{1, 2, \dots, N\}$ be a finite population. Let $\mathbf{y} = (y_1, y_2, \dots, y_m) : U \rightarrow \mathbb{R}^m$ be an \mathbb{R}^m -valued function defined on U (commonly called a “population parameter”). We will use the common notation \mathbf{y}_k for $\mathbf{y}(k)$. We wish to estimate $\mathbf{T}_{\mathbf{y}} := \sum_{k \in U} \mathbf{y}_k \in \mathbb{R}^m$ via survey sampling. Let $p : \mathcal{S} \rightarrow (0, 1]$ be our chosen sampling design, where $\mathcal{S} \subseteq \mathcal{P}(U)$ is the set of all possible samples in the design, and $\mathcal{P}(U)$ is the power set of U .

Proposition 8.1

Suppose the first-order and second-order inclusion probabilities of $p : \mathcal{S} \rightarrow (0, 1]$ are all strictly positive, i.e.

$$\pi_k := \sum_{s \ni k} p(s) = \sum_{k \in U} I_k(s) p(s) > 0 \quad \text{and} \quad \pi_{ik} := \sum_{s \ni i,k} p(s) = \sum_{i,k \in U} I_i(s) I_k(s) p(s) > 0,$$

for any $i, k \in U$. Then, an unbiased estimator for the covariance matrix of the Horvitz-Thompson estimator

$$\widehat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}(s) := \sum_{k \in s} \frac{1}{\pi_k} \mathbf{y}_k$$

is given by:

$$\widehat{\text{Var}}[\widehat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}](s) = \sum_{i,k \in s} \left(\frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}} \right) \cdot \left(\frac{\mathbf{y}_i}{\pi_i} \right) \cdot \left(\frac{\mathbf{y}_k}{\pi_k} \right)^T, \quad \text{for each } s \in \mathcal{S}.$$

PROOF By Proposition 7.5, for any random variable (a.k.a. estimator) $\hat{\mathbf{T}}_{\mathbf{y};w}$ linear in the population parameter $\mathbf{y} : \mathcal{S} \rightarrow \mathbb{R}^m$ with weights $w_k : \mathcal{S} \rightarrow \mathbb{R}$, $k \in U$, the following

$$\widehat{\text{Var}}\left[\hat{\mathbf{T}}_{\mathbf{y};w}\right](s) := \sum_{i,k \in s} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \mathbf{y}_i \cdot \mathbf{y}_k^T \quad (8.1)$$

always gives an unbiased estimator for the covariance matrix of $\hat{\mathbf{T}}_{\mathbf{y};w}$. For the Horvitz-Thompson estimator, the weights are $w_k = 1/\pi_k$, for each $k \in U$, and the weights are independent of the sample $s \in \mathcal{S}$. Thus, for the Horvitz-Thompson estimator, the right-hand side of equation (8.1) becomes:

$$\begin{aligned} \sum_{i,k \in s} \frac{\text{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \mathbf{y}_i \cdot \mathbf{y}_k^T &= \sum_{i,k \in s} \frac{\text{Cov}(I_i, I_k)}{\pi_{ik}} \left(\frac{\mathbf{y}_i}{\pi_i}\right) \cdot \left(\frac{\mathbf{y}_k}{\pi_k}\right)^T \\ &= \sum_{i,k \in s} \frac{E(I_i I_k) - E(I_i)E(I_k)}{\pi_{ik}} \left(\frac{\mathbf{y}_i}{\pi_i}\right) \cdot \left(\frac{\mathbf{y}_k}{\pi_k}\right)^T \\ &= \sum_{i,k \in s} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}} \left(\frac{\mathbf{y}_i}{\pi_i}\right) \cdot \left(\frac{\mathbf{y}_k}{\pi_k}\right)^T, \end{aligned}$$

which coincides with the right-hand side of the equation of the conclusion of the present Proposition. Thus this present Proposition is but a special case of Proposition 7.5, specialized to the Horvitz-Thompson estimator, and the proof is now complete. \square

9 Estimation of Domain Totals

10 Calibrated linear estimators for (multivariate) population totals

Definition 10.1

Let $\hat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S} \rightarrow \mathbb{R}^m$ be an \mathbb{R}^m -valued random variable which is linear in the \mathbb{R}^m -valued population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$, i.e.

$$\begin{aligned} \hat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S} &\rightarrow \mathbb{R}^m \\ s &\mapsto \sum_{k \in s} w_k(s) \cdot \mathbf{y}_k = \sum_{k \in U} I_k(s) w_k(s) \cdot \mathbf{y}_k, \end{aligned}$$

where, for each $k \in U$, $w_k : \mathcal{S} \rightarrow \mathbb{R}$ is itself an \mathbb{R} -valued random variable, and $I_k : \mathcal{S} \rightarrow \{0, 1\}$ is the indicator random variable defined by:

$$I_k(s) = \begin{cases} 1, & \text{if } k \in s, \\ 0, & \text{otherwise} \end{cases}$$

Let $x : U \rightarrow \mathbb{R}$ be an \mathbb{R} -valued population parameter and $T_x := \sum_{k \in U} x_k$.

Then, $\hat{\mathbf{T}}_{\mathbf{y};w}$ is said to be calibrated with respect to x if

$$\sum_{k \in s} w_k(s) x_k = T_x, \text{ for each } s \in \mathcal{S}.$$

Example 10.2

If the sampling design has fixed sample size and each of its first-order inclusion probabilities is strictly positive, then Horvitz-Thompson estimator is calibrated with respect to the first-order inclusion probabilities.

To see this, let $U = \{1, 2, \dots, N\}$ be a finite population, $\mathbf{y} : U \rightarrow \mathbb{R}^m$ a population parameter, and $p : \mathcal{S} \subset \mathcal{P}(U) \rightarrow (0, 1]$ a sampling design such that $\pi_k := \sum_{s \ni k} p(s) > 0$, for each $k \in U$. The Horvitz-Thompson estimator $\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}} : \mathcal{S} \rightarrow \mathbb{R}^m$ is then well-defined and is given by:

$$\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}(s) := \sum_{k \in s} \frac{\mathbf{y}_k}{\pi_k}$$

Let $x : U \rightarrow \mathbb{R}$ be defined by

$$x_k = \pi_k, \text{ for each } k \in U,$$

i.e. x_k is simply the inclusion probability of $k \in U$ under the sampling design $p : \mathcal{S} \rightarrow (0, 1]$.

Now, suppose that the sampling design has a fixed sample size n , and we shall show that $\hat{\mathbf{T}}_{\mathbf{y}}^{\text{HT}}$ is consequently calibrated with respect to $x : U \rightarrow \mathbb{R}$. Indeed, recall that the weights of the Horvitz-Thompson estimator are simply $w_k(s) = 1/\pi_k$, for each $k \in U$ and each $s \in \mathcal{S}$. Hence,

$$\sum_{k \in s} w_k(s) x_k = \sum_{k \in s} \frac{1}{\pi_k} \pi_k = \sum_{k \in s} 1 = \left(\begin{array}{c} \text{sample} \\ \text{size of } s \end{array} \right) = n,$$

since the sampling design has fixed size n . On the other hand,

$$T_x = \sum_{k \in U} x_k = \sum_{k \in U} \pi_k = \sum_{k \in U} E[I_k] = E\left[\sum_{k \in U} I_k\right] = E\left[\begin{array}{c} \text{sample} \\ \text{size} \end{array}\right] = n,$$

again since the sample size is fixed and equals n . Therefore, we have, for any $s \in \mathcal{S}$,

$$\sum_{k \in s} w_k(s) x_k = n = T_x$$

Therefore, the Horvitz-Thompson estimator, under the assumption of fixed sample size, is indeed calibrated with respect to the inclusion probabilities $x : U \rightarrow \mathbb{R}$, $x_k = \pi_k := \sum_{s \ni k} p(s)$, for each $k \in U$. \square

Proposition 10.3

Let $\hat{\mathbf{T}}_{\mathbf{y};w,x} : \mathcal{S} \rightarrow \mathbb{R}^m$ be an \mathbb{R}^m -valued random variable which is linear in the \mathbb{R}^m -valued population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$ and calibrated with respect to the population parameter $x : U \rightarrow \mathbb{R}$, with $x_k \neq 0$ for each $k \in U$.

Then, the mean squared error matrix of $\hat{\mathbf{T}}_{\mathbf{y};w,x}$ as an estimator of $\mathbf{T}_{\mathbf{y}}$ is given by:

$$\text{MSE}\left[\hat{\mathbf{T}}_{\mathbf{y};w,x}\right] = -\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right) \cdot \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^T x_i x_k \in \mathbb{R}^{m \times m}, \text{ where } a_{ik} := E[(I_i w_i - 1)(I_k w_k - 1)].$$

PROOF

$$\begin{aligned} \text{MSE}\left[\hat{\mathbf{T}}_{\mathbf{y};w,x}\right] &= E\left[\left(\hat{\mathbf{T}}_{\mathbf{y};w,x} - \mathbf{T}_{\mathbf{y}}\right) \cdot \left(\hat{\mathbf{T}}_{\mathbf{y};w,x} - \mathbf{T}_{\mathbf{y}}\right)^T\right] = E\left[\left(\sum_{i \in U} (I_i w_i - 1) \mathbf{y}_i\right) \cdot \left(\sum_{k \in U} (I_k w_k - 1) \mathbf{y}_k\right)^T\right] \\ &= \sum_{i \in U} \sum_{k \in U} E[(I_i w_i - 1)(I_k w_k - 1)] \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T = \sum_{k \in U} a_{kk} \cdot \mathbf{y}_k \cdot \mathbf{y}_k^T + \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\ &= \sum_{k \in U} a_{kk} \left(\frac{\mathbf{y}_k \cdot \mathbf{y}_k^T}{x_k^2}\right) x_k^2 + \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left(\frac{\mathbf{y}_i}{x_i}\right) \cdot \left(\frac{\mathbf{y}_k}{x_k}\right)^T x_i x_k \end{aligned}$$

On the other hand,

$$\begin{aligned} &-\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right) \cdot \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^T x_i x_k \\ &= -\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left[\left(\frac{\mathbf{y}_i}{x_i}\right) \left(\frac{\mathbf{y}_i}{x_i}\right)^T - \left(\frac{\mathbf{y}_i}{x_i}\right) \left(\frac{\mathbf{y}_k}{x_k}\right)^T - \left(\frac{\mathbf{y}_k}{x_k}\right) \left(\frac{\mathbf{y}_i}{x_i}\right)^T + \left(\frac{\mathbf{y}_k}{x_k}\right) \left(\frac{\mathbf{y}_k}{x_k}\right)^T\right] x_i x_k \\ &= -\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left[\left(\frac{\mathbf{y}_i}{x_i}\right) \left(\frac{\mathbf{y}_i}{x_i}\right)^T + \left(\frac{\mathbf{y}_k}{x_k}\right) \left(\frac{\mathbf{y}_k}{x_k}\right)^T\right] x_i x_k + \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left(\frac{\mathbf{y}_i}{x_i}\right) \left(\frac{\mathbf{y}_k}{x_k}\right)^T x_i x_k \end{aligned}$$

Thus, the proof of the present Proposition will be complete once we show:

$$\underbrace{\sum_{k \in U} a_{kk} \left(\frac{\mathbf{y}_k}{x_k} \right) \left(\frac{\mathbf{y}_k}{x_k} \right)^T x_k^2}_{\frac{1}{2} \sum_{\substack{i, k \in U \\ i=k}} a_{ik} \left[\left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T + \left(\frac{\mathbf{y}_k}{x_k} \right) \left(\frac{\mathbf{y}_k}{x_k} \right)^T \right] x_i x_k} = -\frac{1}{2} \sum_{\substack{i, k \in U \\ i \neq k}} a_{ik} \left[\left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T + \left(\frac{\mathbf{y}_k}{x_k} \right) \left(\frac{\mathbf{y}_k}{x_k} \right)^T \right] x_i x_k,$$

which is equivalent to:

$$\sum_{i \in U} \sum_{k \in U} a_{ik} \left[\left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T + \left(\frac{\mathbf{y}_k}{x_k} \right) \left(\frac{\mathbf{y}_k}{x_k} \right)^T \right] x_i x_k = 0. \quad (10.2)$$

Observe that

$$\begin{aligned} \text{LHS}(10.2) &= \sum_{i \in U} \sum_{k \in U} a_{ik} \left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T x_i x_k + \sum_{i \in U} \sum_{k \in U} a_{ik} \left(\frac{\mathbf{y}_k}{x_k} \right) \left(\frac{\mathbf{y}_k}{x_k} \right)^T x_i x_k \\ &= 2 \sum_{i \in U} \sum_{k \in U} a_{ik} \left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T x_i x_k = 2 \sum_{i \in U} x_i \left(\frac{\mathbf{y}_i}{x_i} \right) \left(\frac{\mathbf{y}_i}{x_i} \right)^T \left(\sum_{k \in U} a_{ik} x_k \right). \end{aligned}$$

Hence, (10.2) follows once we show

$$\sum_{k \in U} a_{ik} x_k = 0, \quad \text{for each } i \in U. \quad (10.3)$$

Lastly, we now claim that (10.3) follows from the hypothesis that $\hat{T}_{\mathbf{y};w;x}$ is calibrated with respect to x . Indeed,

$$\begin{aligned} \sum_{k \in U} a_{ik} x_k &= \sum_{k \in U} E[(I_i w_i - 1)(I_k w_k - 1)] x_k = \sum_{k \in U} \left[\sum_{s \in \mathcal{S}} p(s) (I_i(s) w_i(s) - 1)(I_k(s) w_k(s) - 1) \right] x_k \\ &= \sum_{s \in \mathcal{S}} p(s) \cdot (I_i(s) w_i(s) - 1) \cdot \left[\sum_{k \in U} (I_k(s) w_k(s) - 1) \cdot x_k \right] \\ &= \sum_{s \in \mathcal{S}} p(s) \cdot (I_i(s) w_i(s) - 1) \cdot \underbrace{\left[\left(\sum_{k \in \mathcal{S}} w_k(s) x_k \right) - T_x \right]}_0 \\ &= 0 \end{aligned}$$

The proof of the present Proposition is now complete. □

Proposition 10.4 (The Yates-Grundy-Sen Variance Estimator for calibrated linear population total estimators)

Let $p : \mathcal{S} \rightarrow (0, 1]$ be a sampling design each of whose first-order and second-order inclusion probabilities is strictly positively. Let $\hat{\mathbf{T}}_{\mathbf{y};w;x} : \mathcal{S} \rightarrow \mathbb{R}^m$ be a random variable which is linear in the population parameter $\mathbf{y} : U \rightarrow \mathbb{R}^m$ and calibrated with respect to the population parameter $x : U \rightarrow \mathbb{R}$, with $x_k \neq 0$ for each $k \in U$. Suppose that $\hat{\mathbf{T}}_{\mathbf{y};w;x}$ is an unbiased estimator for $\mathbf{T}_{\mathbf{y}} := \sum_{k \in U} \mathbf{y}_k$, for arbitrary \mathbf{y} . Then, the following is an unbiased estimator of the variance

$\text{Var}[\hat{\mathbf{T}}_{\mathbf{y};w;x}]$ of $\hat{\mathbf{T}}_{\mathbf{y};w;x}$: For each $s \in \mathcal{S}$ admissible in the sampling design $p : \mathcal{S} \rightarrow (0, 1]$,

$$\widehat{\text{Var}}[\hat{\mathbf{T}}_{\mathbf{y};w;x}](s) := -\frac{1}{2} \sum_{\substack{i, k \in s \\ i \neq k}} \left(w_i(s) w_k(s) - \frac{1}{\pi_{ik}} \right) \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k} \right) \cdot \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k} \right)^T x_i x_k$$

Terminology: $\widehat{\text{Var}}\left[\widehat{\mathbf{T}}_{\mathbf{y};w,x}\right]$ is called the Yates-Grundy-Sen Variance Estimator.

PROOF Since $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ is an unbiased estimator for $\mathbf{T}_{\mathbf{y}}$ by hypothesis, we have $\text{Var}\left[\widehat{\mathbf{T}}_{\mathbf{y};w,x}\right] = \text{MSE}\left[\widehat{\mathbf{T}}_{\mathbf{y};w,x}\right]$. By Proposition 10.3, we thus have:

$$\text{Var}\left[\widehat{\mathbf{T}}_{\mathbf{y};w,x}\right] = -\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} a_{ik} \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^2 x_i x_k, \quad \text{where } a_{ik} := E[(I_i w_i - 1)(I_k w_k - 1)].$$

On the other hand,

$$E\left(\widehat{\text{Var}}\left[\widehat{\mathbf{T}}_{\mathbf{y};w,x}\right]\right) = -\frac{1}{2} \sum_{\substack{i,k \in U \\ i \neq k}} E\left[I_i I_k \left(w_i w_k - \frac{1}{\pi_{ik}}\right)\right] \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right) \cdot \left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^T x_i x_k$$

Thus, it remains only to show:

$$a_{ik} = E\left[I_i I_k \left(w_i w_k - \frac{1}{\pi_{ik}}\right)\right].$$

Now,

$$E\left[I_i I_k \left(w_i w_k - \frac{1}{\pi_{ik}}\right)\right] = E[I_i I_k w_i w_k] - \frac{1}{\pi_{ik}} E[I_i I_k] = E[I_i I_k w_i w_k] - \frac{1}{\pi_{ik}} \pi_{ik} = E[I_i I_k w_i w_k] - 1,$$

and

$$\begin{aligned} a_{ik} &= E[(I_i w_i - 1)(I_k w_k - 1)] = E[I_i I_k w_i w_k] - E[I_i w_i] - E[I_k w_k] + 1 \\ &= E[I_i I_k w_i w_k] - 1 - 1 + 1 = E[I_i I_k w_i w_k] - 1 \\ &= E\left[I_i I_k \left(w_i w_k - \frac{1}{\pi_{ik}}\right)\right], \end{aligned}$$

where third last equality follows from Proposition 7.2 and the unbiasedness hypothesis on $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ as an estimator for $\mathbf{T}_{\mathbf{y}}$. The proof of the present Proposition is now complete. \square

11 Conditional inference in finite-population sampling

In this section, we give a justification for making inference conditional on the observed sample size for sampling designs with random sample size.

Observation (“mixture” of experiments) [see [3], p.15.]

Consider a population \mathcal{U} of 1000 units. We wish to estimate the total T_y of a certain population characteristic $\mathbf{y} = (y_1, y_2, \dots, y_{1000})$. Suppose we use the following two-step sampling scheme:

- Step 1: We first flip a fair coin.
Define the random variable X by letting $X = 1$ if the coin lands heads, and $X = 0$ if it lands tails.
- Step 2: If $X = 1$, we select an SRS from \mathcal{U} of size 100. If $X = 0$, we take a census on all of \mathcal{U} .

Let $\mathcal{S} \subset \mathcal{P}(\mathcal{U})$ denote the probability space of all possible samples induced by the (two-step) sampling design above. Note that $\mathcal{S} = \mathcal{S}_0 \sqcup \mathcal{S}_1$, where $\mathcal{S}_0 = \{\mathcal{U}\}$ and \mathcal{S}_1 is the set of all subsets of \mathcal{U} of size 100. The sampling design is determined by the following probability distribution on \mathcal{S} :

$$P(\mathcal{U}) = \frac{1}{2} \quad \text{and} \quad P(s) = \frac{1}{2 \binom{1000}{100}}, \quad \text{for each } s \in \mathcal{S}_1.$$

Let $\hat{T}_y : \mathcal{S} \rightarrow \mathbb{R}$ denote our chosen estimator for T_y . Then the (unconditional) probability distribution of \hat{T}_y can be “decomposed” as follows:

$$\begin{aligned} P(\hat{T}_y = t \mid \mathbf{y}) &= P(\hat{T}_y = t, X = 0 \mid \mathbf{y}) + P(\hat{T}_y = t, X = 1 \mid \mathbf{y}) \\ &= P(\hat{T}_y = t \mid X = 0, \mathbf{y}) \cdot P(X = 0 \mid \mathbf{y}) + P(\hat{T}_y = t \mid X = 1, \mathbf{y}) \cdot P(X = 1 \mid \mathbf{y}) \\ &= P(\hat{T}_y = t \mid X = 0, \mathbf{y}) \cdot P(X = 0) + P(\hat{T}_y = t \mid X = 1, \mathbf{y}) \cdot P(X = 1), \end{aligned}$$

where the last equality follows because the distribution of X is independent of \mathbf{y} . Suppose the observation we make consists of (\hat{T}_y, X) . The unconditional probability distribution of \hat{T}_y , given by $P(\hat{T}_y = t \mid \mathbf{y})$ above, describes of course the randomness of the estimator \hat{T}_y as induced by both the randomness of the sample $s \in \mathcal{S} = \mathcal{S}_0 \sqcup \mathcal{S}_1$ as well as that of X (the outcome of the coin flip in Step 1). Now, suppose we have indeed carried out the sampling procedure and have obtained an observation of (\hat{T}_y, X) . Suppose it happened that $X = 1$. Hence, we know that the estimate $\hat{T}_y(s)$ we actually obtained was generated from an SRS of size 100 (rather than a census). Note also that the probability distribution of X is independent of \mathbf{y} and the observation of X gives no information about \mathbf{y} . **One school of thought therefore argues that downstream inferences about \mathbf{y} should be carried out using the conditional probability $P(\hat{T}_y = t \mid X = 1, \mathbf{y})$, rather than the unconditional probability $P(\hat{T}_y = t \mid \mathbf{y})$.** In other words, in the present example, as far as making inferences about \mathbf{y} is concerned, only the randomness in Step 2 is relevant, and the randomness in Step 1 (i.e. the randomness of X , the outcome of the coin flip) is irrelevant to any inference about \mathbf{y} . Consequently randomness of X “should” be removed in any inference procedure for \mathbf{y} , and this is achieved by conditioning on the observed value of X . \square

Conditioning on obtained sample size for sample designs with random sample size

Suppose \mathcal{U} is a finite population. We wish to estimate the total $T_y = \sum_{i \in \mathcal{U}} y_i$ of a population characteristic $\mathbf{y} : \mathcal{U} \rightarrow \mathbb{R}$, using a sample design $p : \mathcal{S} \rightarrow [0, 1]$ and an estimator $\hat{T} : \mathcal{S} \rightarrow \mathbb{R}$. **We make the assumption that the sampling design p is independent of \mathbf{y} .** Let $N : \mathcal{S} \rightarrow \mathbb{N} \cup \{0\}$ be the random variable of sample size, i.e. $N(s)$ = number of elements in s , for each possible sample $s \in \mathcal{S}$. Then,

$$\begin{aligned} P(\hat{T} = t \mid \mathbf{y}) &= \sum_n P(\hat{T} = t, N = n \mid \mathbf{y}) \\ &= \sum_n P(\hat{T} = t \mid N = n, \mathbf{y}) \cdot P(N = n \mid \mathbf{y}) \\ &= \sum_n P(\hat{T} = t \mid N = n, \mathbf{y}) \cdot P(N = n), \end{aligned}$$

where the last equality follows from the assumed independence of the probability distribution $p : \mathcal{S} \rightarrow [0, 1]$ (hence that of N) from \mathbf{y} . The key observation to make now is that: **Although the actual sampling procedure operationally may or may not have been a two-step procedure, the independence of p from \mathbf{y} makes it probabilistically equivalent to a two-step procedure, as shown by the above decomposition of $P(\hat{T} = t \mid \mathbf{y})$** — Step (1): randomly select a sample size $N = n$ according to the distribution $P(N = n)$, and then Step (2): randomly select a sample s of size n chosen in Step (1) according to the distribution $P(s \mid N = n)$. By the statistical reasoning explained in the preceding observation, it follows that post-sampling inference about \mathbf{y} should be made based on the conditional distribution $P(\hat{T} = t \mid N = n, \mathbf{y})$, rather than the unconditional distribution $P(\hat{T} = t \mid \mathbf{y})$. This is because the sampling scheme is probabilistically equivalent to a two-step procedure, with the probability distribution of the first step (choosing a sample size) independent of the parameters of interest (T_y), and thus only the probability distribution of the second step (choosing a sample of the size chosen in first step) should be used to make inference about T_y . \square

Caution

In more formal parlance, the random variable $N : \mathcal{S} \rightarrow \mathbb{N} \cup \{0\}$ is ancillary to the parameter \mathbf{y} . Thus, conditioning on sample size, for finite-population sampling schemes with random sample size, *partially* conforms to the **Conditionality**

Principle, which states that statistical inference about a parameter should be made conditioned on observed values of statistics ancillary to that parameter. The conformance is only partial due to the (obvious) fact that it is the sample s itself which is ancillary to the parameter of interest \mathbf{y} , not just its sample size $N(s)$. Thus, full conformance to the Conditionality Principle would require inference about \mathbf{y} be made conditioned on the observed sample s itself (rather than its size $N(s)$). However, if we did condition on the obtained sample s itself, the domain of the estimator \hat{T} would be restricted to the singleton $\{s\}$, and \hat{T} could then attain only one value under conditioning on s , and no randomization-based (i.e. design-based) inference — apart from the observed value of $\hat{T}(s)$ — could be made any longer.

References

- [1] HÁJEK, J. Limiting distributions in simple random sampling from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), 361–374.
- [2] LOHR, S. L. *Sampling: Design and Analysis*, first ed. Duxbury Press, 1999.
- [3] VALLIANT, R., DORFMAN, A. H., AND ROYALL, R. M. *Finite Population Sampling and Inference*, first ed. John-Wiley & Sons, 2000.