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

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

**Definition 1.1** *A* population characteristic *is an $\mathbb{R}$-valued function $y : \mathcal{U} \longrightarrow \mathbb{R}$ 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 $\overline{y}$, of $y$ is defined by:*

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

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

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

In survey sampling, we seek to estimate population total $t$ and population mean $\overline{y}$ of a population characteristic $y : \mathcal{U} \longrightarrow \mathbb{R}$ 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 $n \in \{ 1, 2 \ldots, N \}$ be given. We define the probability space $\Omega_{\mathrm{SRS}}(\mathcal{U}, n)$ 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 $\#(\Omega(\mathcal{U}, n)) = \left(\begin{array}{c} N \\ n \end{array}\right)$. Let $\mathcal{P}(\Omega(\mathcal{U}, n))$ be the power set of $\Omega(\mathcal{U}, n)$. Define $\mu : \Omega \longrightarrow \mathbb{R}$ to be the "uniform" probability measure on the (finite) $\sigma$-algebra $\mathcal{P}(\Omega(\mathcal{U}, n))$ determined by:*

$$\mu(\omega) \;=\; \frac{1}{\left(\begin{array}{c} N \\ n \end{array}\right)} \;=\; \frac{n!(N-n)!}{N!}\,, \quad \text{for each } \omega \in \Omega(\mathcal{U}, n).$$

*Then, $\Omega_{\mathrm{SRS}}(\mathcal{U}, n)$ 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 $\widehat{t}_{\mathrm{SRS}}$ *of the population characteristic $y$ is, by definition, the random variable $\widehat{t}_{\mathrm{SRS}} : \Omega_{\mathrm{SRS}}(\mathcal{U}, n) \longrightarrow \mathbb{R}$ defined by*

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

*The* simple-random-sampling sample mean $\widehat{\overline{y}}_{\mathrm{SRS}}$ *of the population characteristic $y$ is, by definition, the random variable $\widehat{\overline{y}}_{\mathrm{SRS}} : \Omega_{\mathrm{SRS}}(\mathcal{U}, n) \longrightarrow \mathbb{R}$ defined by*

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

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

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

**Proposition 2.3**

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

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*

$$\frac{\widehat{\overline{y}}_{\rm SRS} - \overline{y}}{\sqrt{\left(1-\dfrac{n}{N}\right)\dfrac{S^2}{n}}}$$

*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* $\overline{y}$ *is given by:*

$$\widehat{\overline{y}}_{\rm SRS} \,\pm\, z_{\alpha/2}\cdot\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}$$

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

$$\widehat{\overline{y}}_{\rm SRS} \pm \mathrm{SE}\left[\,\widehat{\overline{y}}_{\rm SRS}\,\right] \;=\; \widehat{\overline{y}}_{\rm SRS} \pm \sqrt{\left(1-\frac{n}{N}\right)\frac{\widehat{s^2}_{\rm SRS}}{n}}$$

*where*

$$\mathrm{SE}\left[\,\widehat{\overline{y}}_{\rm SRS}\,\right] \;:=\; \sqrt{\widehat{\mathrm{Var}}\left[\,\widehat{\overline{y}}_{\rm SRS}\,\right]} \;=\; \sqrt{\left(1-\frac{n}{N}\right)\frac{\widehat{s^2}_{\rm SRS}}{n}}$$

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

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

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

**Immediate observations:**

- $\widehat{t}_{\mathrm{SRS}} = \dfrac{N}{n}\sum_{i=1}^{N} Z_i\,y_i$,, as random variables on $(\Omega, P)$, i.e.

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

- $\widehat{\overline{y}}_{\mathrm{SRS}} = \dfrac{1}{n}\sum_{i=1}^{N} Z_i\,y_i$, as random variables on $(\Omega, P)$, i.e.

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

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

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

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

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

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

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

- For $i\neq j$, we have $E[\,Z_i\cdot Z_j\,] = \left(\dfrac{n-1}{N-1}\right)\cdot\left(\dfrac{n}{N}\right)$. 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 $\mathrm{Cov}\,(Z_i,Z_j) = -\dfrac{1}{N-1}\left(1-\dfrac{n}{N}\right)\left(\dfrac{n}{N}\right) \leq 0$. Indeed,

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

Proof of Proposition 2.3

1.

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

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

2.

$$E\left[\,\widehat{t}_{\mathrm{SRS}}\,\right] \;=\; E\left[\,N\cdot\widehat{\overline{y}}_{\mathrm{SRS}}\,\right] \;=\; N\cdot E\left[\,\widehat{\overline{y}}_{\mathrm{SRS}}\,\right] \;=\; N\cdot\overline{y} \;=\; N\cdot\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) \;=\; \sum_{i=1}^{N} y_i \;=:\; t.$$

$$\mathrm{Var}\left[\,\widehat{t}_{\mathrm{SRS}}\,\right] \;=\; \mathrm{Var}\left[\,N\cdot\widehat{\overline{y}}_{\mathrm{SRS}}\,\right] \;=\; N^2\cdot\mathrm{Var}\left[\,\widehat{\overline{y}}_{\mathrm{SRS}}\,\right] \;=\; N^2\left(1-\frac{n}{N}\right)\frac{S^2}{n}$$

3.
$$\begin{aligned}
E\left[\,\widehat{s^2}_{\mathrm{SRS}}\,\right] &= E\left[\,\frac{1}{n-1}\sum_{i\in\omega}\left(y_i-\widehat{\overline{y}}_{\mathrm{SRS}}\right)^2\,\right] \;=\; \frac{1}{n-1}\,E\left[\,\sum_{i\in\omega}\left((y_i-\overline{y})-(\,\widehat{\overline{y}}_{\mathrm{SRS}}-\overline{y})\right)^2\,\right] \\
&= \frac{1}{n-1}\,E\left[\,\left(\sum_{i\in\omega}(y_i-\overline{y})^2\right)-n\left(\widehat{\overline{y}}_{\mathrm{SRS}}-\overline{y}\right)^2\,\right] \\
&= \frac{1}{n-1}\left\{\,E\left[\,\sum_{i=1}^{N}Z_i(y_i-\overline{y})^2\,\right]-n\,\mathrm{Var}\left[\,\widehat{\overline{y}}_{\mathrm{SRS}}\,\right]\right\} \\
&= \frac{1}{n-1}\left\{\,\sum_{i=1}^{N}E[\,Z_i\,]\,(y_i-\overline{y})^2-n\left(1-\frac{n}{N}\right)\frac{S^2}{n}\,\right\} \\
&= \frac{1}{n-1}\left\{\,\sum_{i=1}^{N}\frac{n}{N}(y_i-\overline{y})^2-\left(1-\frac{n}{N}\right)S^2\,\right\} \\
&= \frac{1}{n-1}\left\{\,\frac{n(N-1)}{N}\,\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\overline{y})^2-\left(1-\frac{n}{N}\right)S^2\,\right\} \\
&= \frac{1}{n-1}\left\{\,\frac{n(N-1)}{N}-\left(1-\frac{n}{N}\right)\,\right\}S^2 \\
&= \frac{1}{n-1}\left\{\,\frac{nN-n-N+n}{N}\,\right\}S^2 \;=\; S^2
\end{aligned}$$

4. Immediate from preceding statements.

5. Immediate from preceding statements. □

# 3 Stratified Simple Random Sampling

Let $\mathcal{U}=\{\,1,2,\ldots,N\,\}$ 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 $\mathcal{U}_1,\mathcal{U}_2,\ldots,\mathcal{U}_H$ is called a *stratum*. Let $N_h:=\#(\mathcal{U}_h)$, for $h=1,2,\ldots,H$. Note that $N_1+N_2+\cdots+N_H=N$.

In *stratified simple random sampling*, an SRS is taken within each stratum $\mathcal{U}_h$, $h=1,2,\ldots,H$. Let $n_h$, $h=1,2,\ldots,H$, be the number elements in the simple random sample taken in the stratum $\mathcal{U}_h$. In other words, a stratified simple random sample $\omega$ of the stratified population $\mathcal{U}=\bigsqcup_{h=1}^{H}\mathcal{U}_h$ has the form:
$$\omega \;=\; \bigsqcup_{h=1}^{H}\omega_h\,,\quad\text{where }\;\omega_h\in\Omega_{\mathrm{SRS}}(\mathcal{U}_h,n_h)\,,\quad\text{for each }\;h=1,2,\ldots,h\,.$$
Note that $n_1+n_2+\cdots+n_H=:n=\#(\omega)$.

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 $y:\mathcal{U}\longrightarrow\mathbb{R}$ be a population characteristic. Define:
$$\widehat{t}_{\mathrm{Str}} \;:=\; \sum_{h=1}^{H}N_h\cdot\widehat{\overline{y}}_{h,\mathrm{SRS}}$$
$$\widehat{\overline{y}}_{\mathrm{Str}} \;:=\; \frac{1}{N}\cdot\widehat{t}_{\mathrm{Str}} \;=\; \sum_{h=1}^{H}\frac{N_h}{N}\cdot\widehat{\overline{y}}_{h,\mathrm{SRS}}$$

Here,

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

is the SRS estimator of

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

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,

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

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

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

And,

$$E\Big[\,\widehat{\overline{y}}_{\mathrm{Str}}\,\Big] \;=\; E\left[\,\frac{1}{N}\cdot\widehat{t}_{\mathrm{Str}}\,\right] \;=\; \frac{1}{N}\,E\big[\,\widehat{t}_{\mathrm{Str}}\,\big] \;=\; \frac{1}{N}\sum_{i=1}^{N} y_i \;=:\; \overline{y}\,.$$

Furthermore,

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

$$\mathrm{Var}\Big[\,\widehat{\overline{y}}_{\mathrm{Str}}\,\Big] \;=\; \mathrm{Var}\left[\,\frac{1}{N}\cdot\widehat{t}_{\mathrm{Str}}\,\right] \;=\; \frac{1}{N^2}\cdot\mathrm{Var}\big[\,\widehat{t}_{\mathrm{Str}}\,\big] \;=\; \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2\left(1-\frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}.$$

**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 $h = 1, 2, \ldots, H$, is chosen such that $n_h/N_h = n/N$. Consequently,

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

where

$$\mathrm{SSW} \;:=\; \sum_{h=1}^{H}\sum_{i\in\mathcal{U}_h}\left(y_i - \overline{y}_{\mathcal{U}_h}\right)^2 \;=\; \sum_{h=1}^{H}(N_h-1)\,S_h^2.$$

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

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

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

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

where

$$\mathrm{SSB} \;:=\; \sum_{h=1}^{H} N_h\left(\overline{y}_{\mathcal{U}_h} - \overline{y}_{\mathcal{U}}\right)^2 \;=\; \sum_{h=1}^{H}\sum_{i \in \mathcal{U}_h}\left(\overline{y}_{\mathcal{U}_h} - \overline{y}_{\mathcal{U}}\right)^2$$

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:

$$\mathrm{SSTO} \;:=\; \sum_{i=1}^{N}\left(y_i - \overline{y}_{\mathcal{U}}\right)^2 \;=\; \sum_{h=1}^{H}\sum_{i \in \mathcal{U}_h}\left(y_i - \overline{y}_{\mathcal{U}}\right)^2 \,.$$

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

$$\sum_{h=1}^{H}\left(1 - \frac{N_h}{N}\right)S_h^2 \;\leq\; \mathrm{SSB} \quad\Longrightarrow\quad \mathrm{Var}\big[\,\widehat{t}_{\mathrm{PropStr}}\,\big] \leq \mathrm{Var}\big[\,\widehat{t}_{\mathrm{SRS}}\,\big] \,.$$

*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 $\mathcal{U} = \displaystyle\bigsqcup_{i=1}^{N}\mathcal{C}_i$ 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) $\omega_1 = \{\,\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \ldots, \mathcal{C}_{i_n}\,\}$ 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 $\omega := \displaystyle\bigsqcup_{\mathcal{C} \in \omega_1}\omega_{\mathcal{C}}$. In other words, the sample $\omega$ 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 $\widehat{t}_{\mathrm{HT}}$, as defined below, is an unbiased estimator for the total of an $\mathbb{R}$-valued population characteristic $y : \mathcal{U} \longrightarrow \mathbb{R}$.

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

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 $\widehat{t}_{\mathrm{HT}}$ is given by:

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

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 $\mathrm{Var}\left(\widehat{t}_{\mathrm{HT}}\right)$ 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 $i = 1, 2, \ldots, N$.

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

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

# 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 $\omega_1 = \{\,\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_N\,\}$. In particular, $n = N$.

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

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

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

# 7 Linear estimators for population totals

Let $U = \{1, 2, \ldots, N\}$ be a finite population. Let $y : U \longrightarrow \mathbb{R}$ be an $\mathbb{R}$-valued function defined on $U$ (commonly called a "population parameter"). We will use the common notation $y_k$ for $y(k)$. We wish to estimate $T_y := \sum_{k \in U} y_k$ via survey sampling. Let $p : \mathcal{S} \longrightarrow (0, 1]$ 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 $\widehat{T}_y : \mathcal{S} \longrightarrow \mathbb{R}$ is said to be* <u>linear *in the population parameter* $y : U \longrightarrow \mathbb{R}$</u> *if it has the following form:*

$$\begin{array}{cccc} \widehat{T}_y : & \mathcal{S} & \longrightarrow & \mathbb{R} \\ & s & \longmapsto & \displaystyle\sum_{k \in s} w_k(s)\, y_k, \end{array}$$

*where, for each $k \in U$, $w_k : \mathcal{S} \longrightarrow \mathbb{R}$ is itself an $\mathbb{R}$-valued random variable. We call the $w_k$'s the* <u>weights</u> *of $\widehat{T}_y$.*

**Nomenclature** In the context of finite-population probability sampling, under a design $p : \mathcal{S} \longrightarrow (0, 1]$, 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 $\widehat{T}_{y;w} : \mathcal{S} \longrightarrow \mathbb{R}$, with $\widehat{T}_{y;w}(s) = \sum_{k \in s} w_k(s)\, y_k$, be a random variable linear in the population parameter $y : U \longrightarrow \mathbb{R}$. Then,*

$$E\left[\, \widehat{T}_{y;w} \,\right] \;=\; T_y, \text{ for arbitrary } y \quad \Longleftrightarrow \quad \sum_{s \ni k} p(s)\, w_k(s) \;=\; 1, \text{ for each } k \in U.$$

PROOF For each $k \in U$, define the indicator random variable $I_k : \mathcal{S} \longrightarrow \{0, 1\}$ by:

$$I_k(s) \;=\; \left\{ \begin{array}{ll} 1\,, & \text{if } k \in s \\ 0\,, & \text{otherwise} \end{array} \right. .$$

Note that

$$E\left[\, \widehat{T}_{y;w} \,\right] \;=\; E\left[ \sum_{k \in s} w_k\, y_k \right] \;=\; E\left[ \sum_{k \in U} I_k w_k\, y_k \right] \;=\; \sum_{k \in U} E[\, I_k w_k \,]\; y_k$$

Hence, since $y : U \longrightarrow \mathbb{R}$ is arbitrary,

$$E\left[\, \widehat{T}_{y;w} \,\right] \;=\; T_y \;:=\; \sum_{k \in U} y_k \quad \Longleftrightarrow \quad \sum_{k \in U} (E[\, I_k w_k \,] - 1) \cdot y_k \;=\; 0 \quad \Longleftrightarrow \quad E[\, I_k w_k \,] \;=\; 1, \text{ for each } k \in U$$

Lastly, note that

$$E[\, I_k w_k \,] \;=\; \sum_{s \in \mathcal{S}} p(s)\, I_k(s)\, w_k(s) \;=\; \sum_{k \ni s} p(s)\, w_k(s),$$

and the proof of the Proposition is complete. □

**Corollary 7.3**

*For a sampling design each of whose inclusion probabilities is strictly positive, the Horvitz-Thompson estimator $\widehat{T}_y^{\mathrm{HT}}$ is well-defined and it is the unique unbiased estimator for $T_y$, for arbitrary $y$, which is linear in $y$ and whose weights are constants in $s$.*

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

$$\widehat{T}_y^{\mathrm{HT}}(s) \;:=\; \sum_{k \in s} \frac{y_k}{\pi_k},$$

where $\pi_k := \sum_{s \ni k} p(s)$ is the inclusion probability of $k \in U$ under the sampling design $p : \mathcal{S} \longrightarrow (0,1]$. Clearly, $\widehat{T}_y^{\mathrm{HT}}$ is linear in $y$ with weights constant in $s$. Next, note that:

$$\begin{aligned}
E\left[\, \widehat{T}_y^{\mathrm{HT}} \,\right] &= E\left[\sum_{k\in s} \frac{y_k}{\pi_k}\right] = E\left[\sum_{k\in U} I_k \frac{y_k}{\pi_k}\right] = \sum_{k\in U} E[\, I_k \,] \frac{y_k}{\pi_k} = \sum_{k\in U}\left[\sum_{s\in\mathcal{S}} p(s) I_k(s)\right]\frac{y_k}{\pi_k} \\
&= \sum_{k\in U}\left[\sum_{s\ni k} p(s)\right]\frac{y_k}{\pi_k} = \sum_{k\in U} \pi_k \frac{y_k}{\pi_k} = \sum_{k\in U} y_k = T_y
\end{aligned}$$

Hence, $\widehat{T}_y^{\mathrm{HT}}$ is an unbiased estimator for $T_y$. Conversely, let

$$\widehat{T}_{y;w}(s) = \sum_{k\in s} w_k\, y_k$$

be any unbiased estimator for $T_y$ which linear in $y$ with weights constant in $s$. Thus,

$$\sum_{k\in U} y_k = T_y = E\left[\, \widehat{T}_{y;w} \,\right] = E\left[\sum_{k\in s} w_k\, y_k\right] = E\left[\sum_{k\in U} I_k\, w_k\, y_k\right] = \sum_{k\in U} E[\, I_k \,]\, w_k\, y_k = \sum_{k\in U} \pi_k\, w_k\, y_k.$$

Since $y$ is arbitrary, the above equation immediately implies that

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

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

**Proposition 7.4**

*Let $\widehat{T}_{y;w} : \mathcal{S} \longrightarrow \mathbb{R}$, with $\widehat{T}_{y;w}(s) = \sum_{k\in s} w_k(s)\, y_k$, be a random variable linear in the population parameter $y : U \longrightarrow \mathbb{R}$. Then, the variance of $\widehat{T}_{y;w}$ is given by:*

$$\mathrm{Var}\left[\, \widehat{T}_{y;w} \,\right] = \sum_{i\in U} a_i\, y_i^2 + \sum_{\substack{i,j\in U \\ i\neq j}} a_{ij}\, y_i\, y_j,$$

*where*

$$a_i := \mathrm{Var}[\, I_i\, w_i \,], \quad a_{ij} := \mathrm{Cov}[\, I_i w_i \,,\, I_j w_j \,], \quad \text{and} \quad I_i(s) = \begin{cases} 1\,, & \text{if } i \in s \\ 0\,, & \text{otherwise} \end{cases}.$$

*Furthermore, if $\widehat{T}_{y;w}$ is an unbiased estimator for $T_y := \sum_{k\in U} y_k$ for arbitrary $y$, then*

$$a_i = \mathrm{Var}[\, I_i w_i \,] = \left(\sum_{s\in\mathcal{S}} p(s)\, I_i(s)\, w_i^2(s)\right) - 1$$

PROOF

$$\begin{aligned}
\mathrm{Var}\left[\, \widehat{T}_{y;w} \,\right] &= \mathrm{Cov}\left[\sum_{i\in U} I_i\, w_i\, y_i \,,\, \sum_{k\in U} I_k\, w_k\, y_k\right] = \sum_{i\in U}\sum_{k\in U} \mathrm{Cov}[\, I_i\, w_i \,,\, I_k\, w_k \,]\, y_i\, y_k \\
&= \sum_{k\in U} \mathrm{Var}[\, I_k\, w_k \,]\, y_k^2 + \sum_{\substack{i,k\in U \\ i\neq k}} \mathrm{Cov}[\, I_i\, w_i \,,\, I_k\, w_k \,]\, y_i\, y_k
\end{aligned}$$

Furthermore, if $\widehat{T}_{y;w}$ is an unbiased estimator for $T_y := \sum_{k\in U} y_k$ for arbitrary $y$, then $\sum_{s\ni k} p(s)\,w_k(s) = 1$, for each $k \in U$, by the preceding Proposition. Hence, for each $k \in U$,

$$\begin{aligned}
a_k \;&:=\; \mathrm{Var}[\, I_k\, w_k \,] \;=\; E[\, I_k^2\, w_k^2 \,] - E[\, I_k\, w_k \,] \\
&= \left(\sum_{s\in\mathcal{S}} p(s)\, I_i^2(s)\, w_i^2(s)\right) - \left(\sum_{s\in\mathcal{S}} p(s)\, I_i(s)\, w_i(s)\right) \;=\; \left(\sum_{s\in\mathcal{S}} p(s)\, I_i^2(s)\, w_i^2(s)\right) - \left(\sum_{s\ni k} p(s)\, w_i(s)\right) \\
&= \left(\sum_{s\in\mathcal{S}} p(s)\, I_i(s)\, w_i^2(s)\right) - 1,
\end{aligned}$$

and the proof of the Proposition is complete. □

**Definition 7.5**

*Let $\widehat{T}_{y;w} : \mathcal{S} \longrightarrow \mathbb{R}$ be a random variable which is linear in the population parameter $y : U \longrightarrow \mathbb{R}$, i.e.*

$$\begin{array}{rccl}
\widehat{T}_{y;w} : & \mathcal{S} & \longrightarrow & \mathbb{R} \\
& s & \longmapsto & \sum_{k\in s} w_k(s)\, y_k,
\end{array}$$

*where, for each $k \in U$, $w_k : \mathcal{S} \longrightarrow \mathbb{R}$ is itself an $\mathbb{R}$-valued random variable. Let $x : U \longrightarrow \mathbb{R}$ be another population parameter and $T_x := \sum_{k\in U} x_k$. Then, $\widehat{T}_{y;w}$ is said to be* <u>calibrated with respect to $x$</u> *if*

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

**Proposition 7.6**

*Let $\widehat{T}_{y;w,x} : \mathcal{S} \longrightarrow \mathbb{R}$ be a random variable which is linear in the population parameter $y : U \longrightarrow \mathbb{R}$ and calibrated with respect to the population parameter $x : U \longrightarrow \mathbb{R}$, with $x_k > 0$ for each $k \in U$. Then, the mean squared error of $\widehat{T}_{y;w,x}$ as an estimator of $T_y$ is given by:*

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

Proof

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

On the other hand,

$$\begin{aligned}
& -\frac{1}{2} \sum_{\substack{i,k\in U \\ i\neq k}} a_{ik} \left(\frac{y_i}{x_i} - \frac{y_k}{x_k}\right)^2 x_i\, x_k \\
=\; & -\frac{1}{2} \sum_{\substack{i,k\in U \\ i\neq k}} a_{ik} \left[\left(\frac{y_i}{x_i}\right)^2 - 2\left(\frac{y_i}{x_i}\right)\left(\frac{y_k}{x_k}\right) + \left(\frac{y_k}{x_k}\right)^2\right] x_i\, x_k \\
=\; & -\frac{1}{2} \sum_{\substack{i,k\in U \\ i\neq k}} a_{ik} \left[\left(\frac{y_i}{x_i}\right)^2 + \left(\frac{y_k}{x_k}\right)^2\right] x_i\, x_k + \sum_{\substack{i,k\in U \\ i\neq k}} a_{ik} \left(\frac{y_i}{x_i}\right) \left(\frac{y_k}{x_k}\right) x_i\, x_k
\end{aligned}$$

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

$$\sum_{k\in U} a_{kk}\left(\frac{y_k}{x_k}\right)^2 x_k^2 \;\;=\;\; -\frac{1}{2}\sum_{\substack{i,k\in U\\ i\neq k}} a_{ik}\left[\left(\frac{y_i}{x_i}\right)^2+\left(\frac{y_k}{x_k}\right)^2\right] x_i\,x_k,$$

which is equivalent to:

$$\sum_{i\in U}\sum_{k\in U} a_{ik}\left[\left(\frac{y_i}{x_i}\right)^2+\left(\frac{y_k}{x_k}\right)^2\right] x_i\,x_k \;\;=\;\; 0. \tag{7.1}$$

Observe that

$$\begin{aligned}
\text{LHS}(7.1) &= \sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{y_i}{x_i}\right)^2 x_i\,x_k + \sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{y_k}{x_k}\right)^2 x_i\,x_k \;\;=\;\; 2\sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{y_i}{x_i}\right)^2 x_i\,x_k \\
&= 2\sum_{i\in U} x_i\left(\frac{y_i}{x_i}\right)^2\left(\sum_{k\in U} a_{ik}x_k\right).
\end{aligned}$$

Hence, (7.1) follows once we show

$$\sum_{k\in U} a_{ik}x_k = 0, \quad \text{for each } i\in U. \tag{7.2}$$

Lastly, we now claim that (7.2) follows from the hypothesis that $\widehat{T}_{y;w;x}$ is calibrated with respect to $x$. Indeed,

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

The proof of the present Proposition is now complete. □

**Proposition 7.7 (The Yates-Grundy-Sen Variance Estimator)**

*Let $p:\mathcal{S}\longrightarrow(0,1]$ be a sampling design each of whose first-order and second-order inclusion probabilities is strictly positively. Let $\widehat{T}_{y;w,x}:\mathcal{S}\longrightarrow\mathbb{R}$ be a random variable which is linear in the population parameter $y:U\longrightarrow\mathbb{R}$ and calibrated with respect to the population parameter $x:U\longrightarrow\mathbb{R}$, with $x_k>0$ for each $k\in U$. Suppose that $\widehat{T}_{y;w,x}$ is an unbiased estimator for $T_y := \sum_{k\in U} y_k$, for arbitrary $y$. Then, the following is an unbiased estimator of the variance $\text{Var}\left[\,\widehat{T}_{y;w,x}\,\right]$ of $\widehat{T}_{y;w,x}$: For each $s\in\mathcal{S}$ admissible in the sampling design $p:\mathcal{S}\longrightarrow(0,1]$,*

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

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

PROOF Since $\widehat{T}_{y;w,x}$ is an unbiased estimator for $T_y$ by hypothesis, we have $\mathrm{Var}\left[\,\widehat{T}_{y;w,x}\,\right] = \mathrm{MSE}\left[\,\widehat{T}_{y;w,x}\,\right]$. By Proposition 7.6, we thus have:

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

On the other hand,

$$E\Big(\,\widehat{\mathrm{Var}}\left[\,\widehat{T}_{y;w,x}\,\right]\Big) \;=\; -\frac{1}{2}\sum_{\substack{i,k\in U\\ i\neq k}} E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right]\left(\frac{y_i}{x_i}-\frac{y_k}{x_k}\right)^2 x_i\,x_k$$

Thus, it remains only to show:

$$a_{ik} \;=\; E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right].$$

Now,

$$E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right] \;=\; E[\,I_iI_kw_iw_k\,]-\frac{1}{\pi_{ik}}E[\,I_iI_k\,] \;=\; E[\,I_iI_kw_iw_k\,]-\frac{1}{\pi_{ik}}\,\pi_{ik} \;=\; E[\,I_iI_kw_iw_k\,]-1,$$

and

$$\begin{aligned}
a_{ik} &= E[\,(I_i\,w_i-1)\,(I_k\,w_k-1)\,] \;=\; E[\,I_i\,I_k\,w_i\,w_k\,]-E[\,I_i\,w_i\,]-E[\,I_k\,w_k\,]+1\\
&= E[\,I_i\,I_k\,w_i\,w_k\,]-\sum_{s\ni i}p(s)\,w_i-\sum_{s\ni k}p(s)\,w_k+1 \;=\; E[\,I_i\,I_k\,w_i\,w_k\,]-1-1+1 \;=\; E[\,I_i\,I_k\,w_i\,w_k\,]-1\\
&= E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right],
\end{aligned}$$

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

# 8 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 $\mathbf{y} = (y_1, y_2, \ldots, y_{1000})$. 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 $\mathcal{S}$:

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

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

$$\begin{aligned}
P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\right) &= P\left(\,\widehat{T}_y = t,\, X = 0 \,\middle|\, \mathbf{y}\right) + P\left(\,\widehat{T}_y = t,\, X = 1 \,\middle|\, \mathbf{y}\right) \\
&= P\left(\,\widehat{T}_y = t \,\middle|\, X = 0,\, \mathbf{y}\right) \cdot P\left(\,X = 0 \,|\, \mathbf{y}\,\right) + P\left(\,\widehat{T}_y = t \,\middle|\, X = 1,\, \mathbf{y}\right) \cdot P\left(\,X = 1 \,|\, \mathbf{y}\,\right) \\
&= P\left(\,\widehat{T}_y = t \,\middle|\, X = 0,\, \mathbf{y}\right) \cdot P\left(\,X = 0\,\right) + P\left(\,\widehat{T}_y = t \,\middle|\, X = 1,\, \mathbf{y}\right) \cdot P\left(\,X = 1\,\right),
\end{aligned}$$

where the last equality follows because the distribution of $X$ is independent of $\mathbf{y}$. Suppose the observation we make consists of $\left(\,\widehat{T}_y\,,\, X\,\right)$. The unconditional probability distribution of $\widehat{T}_y$, given by $P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\right)$ above, describes of course the randomness of the estimator $\widehat{T}_y$ 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 $\left(\,\widehat{T}_y\,,\, X\,\right)$. Suppose it happened that $X = 1$. Hence, we know that the estimate $\widehat{T}_y(s)$ 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 $P\left(\,\widehat{T}_y = t \,\middle|\, X = 1\,,\, \mathbf{y}\right)$, rather than the unconditional probability $P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\right)$. 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 $X$. □

**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 $\mathbf{y} : \mathcal{U} \longrightarrow \mathbb{R}$, using a sample design $p : \mathcal{S} \longrightarrow [\,0\,,1\,]$ and a estimator $\widehat{T} : \mathcal{S} \longrightarrow \mathbb{R}$. We make the assumption that the sampling design $p$ is independent of $\mathbf{y}$. Let $N : \mathcal{S} \longrightarrow \mathbb{N} \cup \{\,0\,\}$ 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,

$$\begin{aligned}
P\left(\,\widehat{T} = t \,\middle|\, \mathbf{y}\right) &= \sum_n P\left(\,\widehat{T} = t,\, N = n \,\middle|\, \mathbf{y}\right) \\
&= \sum_n P\left(\,\widehat{T} = t \,\middle|\, N = n,\, \mathbf{y}\right) \cdot P\left(\,N = n \,|\, \mathbf{y}\,\right) \\
&= \sum_n P\left(\,\widehat{T} = t \,\middle|\, N = n,\, \mathbf{y}\right) \cdot P\left(\,N = n\,\right),
\end{aligned}$$

where the last equality follows from the assumed independence of the probability distribution $p : \mathcal{S} \longrightarrow [\,0\,,\,1\,]$ (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 $P\left(\,\widehat{T} = t \,\middle|\, \mathbf{y}\right)$ — 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 $P(\,s \,|\, N = n\,)$. 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 $P\left(\,\widehat{T} = t \,\middle|\, N = n\,,\, \mathbf{y}\right)$, rather than the unconditional distribution $P\left(\,\widehat{T} = t \,\middle|\, \mathbf{y}\right)$. 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 $T_y$. □

**Caution**

In more formal parlance, the random variable $N : \mathcal{S} \longrightarrow \mathbb{N} \cup \{\,0\,\}$ 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 $\widehat{T}$ would be restricted to the singleton $\{\,s\,\}$, and $\widehat{T}$ 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 $\widehat{T}(s)$ — 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.