Survey samples from a finite population

Let $\mathcal{P} = \{1, \ldots, N\}$ be a finite population of interest.
For each $i \in \mathcal{P}$, let $y_i$ be a value of interest associated with unit $i$.
Let $s = \{i_1, \ldots, i_n\}$ be a subset of distinct elements of $\mathcal{P}$, which is a sample selected with selection probability $p(s)$, where $p$ is known.
The value $y_i$ is observed if and only if $i \in s$.
$Y = \sum_{j=1}^{N} y_j$ is the unknown population total of interest.

Define

$$\pi_i = \text{probability that } i \in s, \quad i = 1, \ldots, N.$$ 

Horvitz-Thompson estimators

- It gives a general method to obtain unbiased estimators.
- All we need is the inclusion probability $\pi_i$, which is known in sample surveys since $p(s)$ is known.
Theorem 3.15.

(i) (Horvitz-Thompson). If \( \pi_i > 0 \) for \( i = 1, \ldots, N \) and \( \pi_i \) is known when \( i \in s \), then \( \hat{Y}_{ht} = \sum_{i \in s} y_i / \pi_i \) is an unbiased estimator of the population total \( Y \).

(ii) Define

\[
\pi_{ij} = \text{probability that } i \in s \text{ and } j \in s, \quad i = 1, \ldots, N, j = 1, \ldots, N.
\]

Then

\[
\text{Var}(\hat{Y}_{ht}) = \sum_{i=1}^{N} \frac{1 - \pi_i}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j \tag{1}
\]

\[
= \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \tag{2}
\]

- Horvitz-Thompson’s idea: inverse probability weighting
- The unbiasedness of the sample mean under simple random sampling without replacement is a special case of Theorem 3.15.
- Extension: \( \mathcal{D} \) is a sample of size \( N \) and \( y_i \) is missing if \( i \not\in s \)
- If \( \pi_j \) is unknown, we need to replace it by an estimator.
Proof.

(i) Let $a_i = 1$ if $i \in s$ and $a_i = 0$ if $i \notin s$, $i = 1, \ldots, N$. Then $E(a_i) = \pi_i$ and

$$E(\hat{Y}_{ht}) = E \left( \sum_{i=1}^{N} \frac{a_i y_i}{\pi_i} \right) = \sum_{i=1}^{N} y_i = Y.$$

(ii) Since $a_i^2 = a_i$,

$$\text{Var}(a_i) = E(a_i) - [E(a_i)]^2 = \pi_i (1 - \pi_i).$$

$$\text{Cov}(a_i, a_j) = E(a_i a_j) - E(a_i)E(a_j) = \pi_{ij} - \pi_i \pi_j, \quad i \neq j.$$

Then

$$\text{Var}(\hat{Y}_{ht}) = \text{Var} \left( \sum_{i=1}^{N} \frac{a_i y_i}{\pi_i} \right)$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2} \text{Var}(a_i) + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{y_i y_j}{\pi_i \pi_j} \text{Cov}(a_i, a_j)$$

$$= \sum_{i=1}^{N} \frac{1 - \pi_i}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j.$$
Proof (continued)

Hence (1) follows.

To show (2), note that

$$ \sum_{i=1}^{N} \pi_i = n \quad \text{and} \quad \sum_{j=1, \ldots, N, j \neq i} \pi_{ij} = (n - 1) \pi_i, $$

which implies

$$ \sum_{j=1, \ldots, N, j \neq i} (\pi_{ij} - \pi_i \pi_j) = (n - 1) \pi_i - \pi_i (n - \pi_i) = -\pi_i (1 - \pi_i). $$

Hence

$$ \sum_{i=1}^{N} \frac{1 - \pi_i}{\pi_i} y_i^2 = \sum_{i=1}^{N} \sum_{j=1, \ldots, N, j \neq i} (\pi_i \pi_j - \pi_{ij}) \frac{y_i^2}{\pi_i^2} $$

$$ = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} \right) $$

and, (2) follows from (1).
How do we get an unbiased estimator of $\text{Var}(\hat{Y}_{ht})$?

Using Horvitz-Thompson’s idea, the following estimators are unbiased:

$$v_1 = \sum_{i \in s} \frac{1 - \pi_i}{\pi_i^2} y_i^2 + 2 \sum_{i \in s} \sum_{j \in s, j > i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j$$

$$v_2 = \sum_{i \in s} \sum_{j \in s, j > i} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2.$$ 

Simple random sampling without replacement

For simple random sampling without replacement,

$$\pi_i = E(a_i) = P(a_i = 1) = \frac{(N-1)}{\binom{N}{n}} = \frac{n}{N}$$

$$\pi_{ij} = E(a_i a_j) = P(a_i = 1, a_j = 1) = \frac{(N-2)}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

$$\hat{Y}_{ht} = \frac{N}{n} \sum_{i \in s} y_i = \frac{N}{n} \sum_{i=1}^{N} a_i y_i = N(\text{the sample mean})$$
\[
\text{Var}(\hat{Y}_{ht}) = \sum_{i=1}^{N} \frac{1 - \frac{n}{N}}{\frac{n}{N}} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} y_i y_j
\]

\[
= \frac{N-n}{n} \sum_{i=1}^{N} y_i^2 - \frac{2(N-n)}{n(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} y_i y_j
\]

\[
= \frac{N-n}{n} \left[ \sum_{i=1}^{N} y_i^2 - \frac{1}{N-1} \sum_{i \neq j} y_i y_j \right]
\]

\[
= \frac{N-n}{n} \frac{N}{N-1} \sum_{i=1}^{N} \left( y_i - \frac{Y}{N} \right)^2
\]

\[
= \left( 1 - \frac{n}{N} \right) \frac{N^2}{n} \frac{1}{N-1} \sum_{i=1}^{N} \left( y_i - \frac{Y}{N} \right)^2
\]

\[
= N^2 \left( 1 - \frac{n}{N} \right) \frac{S^2}{n}
\]

\(\frac{n}{N}\) is called the finite sample fraction and \(1 - \frac{n}{N}\) is called the finite sample correction.

\(S^2\) is called the population variance.
For simple random sampling without replacement, variance estimators $\nu_1$ and $\nu_2$ are the same.

Note that

\[
\nu_1 = \sum_{i \in s} \left( \frac{1}{n} \frac{n}{N} - \frac{n^2}{N^2} \right) y_i^2 + 2 \sum_{i \in s} \sum_{j \in s, j > i} \frac{1}{n(n-1)} \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} y_i y_j
\]

\[
= \frac{N(N - n)}{n^2} \sum_{i \in s} y_i^2 - \frac{N(N - n)}{n^2(n-1)} \sum_{i, j \in s, i \neq j} y_i y_j
\]

\[
= \frac{N(N - n)}{n(n-1)} \sum_{i \in s} y_i^2 - \frac{N(N - n)}{n^2(n-1)} \sum_{i, j \in s} y_i y_j
\]

\[
= \frac{N(N - n)}{n(n-1)} \left[ \sum_{i \in s} y_i^2 - \frac{1}{n} \left( \sum_{i \in s} y_i \right)^2 \right]
\]

\[
= N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n}
\]
where
\[ s^2 = \frac{1}{n-1} \sum_{i \in s} (y_i - \bar{y})^2 \]
is called the sample variance.

Since \( E(v_1) = \text{Var}(\hat{Y}_{ht}) \), we have shown that \( E(s^2) = S^2 \).

Since \( s^2 \) is symmetric in its arguments, the early result implies that \( s^2 \) is a UMVUE of \( S^2 \) and \( v_1 \) is a UMVUE of \( \text{Var}(\hat{Y}_{ht}) \), under simple random sampling without replacement.

To finish, we note that
\[
v_2 = v_1 + \sum_{i,j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_i^2}{\pi_i^2} \]
\[
= v_1 + \sum_{i,j \in s} \left( \frac{1}{\pi_i^2} - \frac{1}{\pi_{ij}} \right) y_i^2
\]
\[
= v_1 + \frac{N^2}{n} \sum_{i \in s} y_i^2 - \frac{N}{n} \sum_{i \in s} y_i^2 - \frac{N(N-1)}{n(n-1)} \sum_{i,j \in s, j \neq i} y_i^2
\]
\[
= v_1
\]
Deriving asymptotically unbiased estimators

An exactly unbiased estimator may not exist, or is hard to obtain. We often derive asymptotically unbiased estimators. Functions of sample means are popular estimators.

Functions of unbiased estimators

If the parameter to be estimated is $\vartheta = g(\theta)$ with a vector-valued parameter $\theta$ and $U_n$ is a vector of unbiased estimators of components of $\theta$, then $T_n = g(U_n)$ is often asymptotically unbiased for $\vartheta$.

Note that $E(T_n) = Eh(U_n)$ may not exist.

Assume that $g$ is differentiable and

$$c_n(U_n - \theta) \rightarrow_d Y.$$

Then, by Theorem 2.6,

$$\text{amse}_{T_n}(P) = E\{[\nabla g(\theta)]^\tau Y\}^2 / c_n^2$$

Hence, $T_n$ has a good performance in terms of amse if $U_n$ is optimal in terms of mse (such as the UMVUE or BLUE).
Method of moments

The method of moments is the oldest method of deriving asymptotically unbiased estimators, which may not be the best estimators, but they are simple and can be used as initial estimators.

Consider a parametric problem where $X_1, \ldots, X_n$ are i.i.d. random variables from $P_\theta$, $\theta \in \Theta \subset \mathbb{R}^k$, and $E|X_1|^k < \infty$.

Let $\mu_j = EX_1^j$ be the $j$th moment of $P$ and let

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$$

be the $j$th sample moment, which is an unbiased estimator of $\mu_j$, $j = 1, \ldots, k$.

Typically,

$$\mu_j = h_j(\theta), \quad j = 1, \ldots, k,$$

for some functions $h_j$ on $\mathbb{R}^k$. 


By substituting $\mu_j$’s on the left-hand side of (3) by the sample moments $\hat{\mu}_j$, we obtain a moment estimator $\hat{\theta}$, i.e., $\hat{\theta}$ satisfies

$$\hat{\mu}_j = h_j(\hat{\theta}), \quad j = 1, \ldots, k,$$

which is a sample analogue of (3).

This method of deriving estimators is called the method of moments. An important statistical principle, the substitution principle, is applied in this method.

Let $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_k)$ and $h = (h_1, \ldots, h_k)$. Then $\hat{\mu} = h(\hat{\theta})$.

If the inverse function $h^{-1}$ exists, then the unique moment estimator of $\theta$ is $\hat{\theta} = h^{-1}(\hat{\mu})$.

When $h^{-1}$ does not exist (i.e., $h$ is not one-to-one), any solution of $\hat{\mu} = h(\hat{\theta})$ is a moment estimator of $\theta$.

If possible, we always choose a solution $\hat{\theta}$ in the parameter space $\Theta$. 

In some cases, however, a moment estimator does not exist (see Exercise 111).

Moment estimators may not be unique.

We usually use moments with the lowest possible order.

Assume that \( \hat{\theta} = g(\hat{\mu}) \) for a function \( g \).

If \( h^{-1} \) exists, then \( g = h^{-1} \).

If \( g \) is continuous at \( \mu = (\mu_1, \ldots, \mu_k) \), then \( \hat{\theta} \) is strongly consistent for \( \theta \), since \( \hat{\mu}_j \rightarrow_{a.s.} \mu_j \) by the SLLN.

If \( g \) is differentiable at \( \mu \) and \( E|X_1|^{2k} < \infty \), then \( \hat{\theta} \) is asymptotically normal, by the CLT and Theorem 1.12, and

\[
\text{amse}_{\hat{\theta}}(\theta) = n^{-1}[\nabla g(\mu)]^\tau V_\mu \nabla g(\mu),
\]

where \( V_\mu \) is a \( k \times k \) matrix whose \((i,j)\)th element is \( \mu_{i+j} - \mu_i \mu_j \).

Furthermore, the \( n^{-1} \) order asymptotic bias of \( \hat{\theta} \) is

\[
(2n)^{-1}\text{tr}\left(\nabla^2 g(\mu) V_\mu\right).
\]
Example 3.24

Let $X_1, \ldots, X_n$ be i.i.d. from a population $P_\theta$ indexed by the parameter $	heta = (\mu, \sigma^2)$, where $\mu = \mathbb{E}X_1 \in \mathbb{R}$ and $\sigma^2 = \text{Var}(X_1) \in (0, \infty)$. This includes cases such as the family of normal distributions, double exponential distributions, or logistic distributions (Table 1.2, page 20).

Since $\mathbb{E}X_1 = \mu$ and $\mathbb{E}X_1^2 = \text{Var}(X_1) + (\mathbb{E}X_1)^2 = \sigma^2 + \mu^2$, setting $\hat{\mu}_1 = \mu$ and $\hat{\mu}_2 = \sigma^2 + \mu^2$ we obtain the moment estimator

$$\hat{\theta} = \left( \bar{X}, \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right) = \left( \bar{X}, \frac{n-1}{n} S^2 \right).$$

Note that $\bar{X}$ is unbiased, but $\frac{n-1}{n} S^2$ is not.

If $X_i$ is normal, then $\hat{\theta}$ is sufficient and is nearly the same as an optimal estimator such as the UMVUE.

On the other hand, if $X_i$ is from a double exponential or logistic distribution, then $\hat{\theta}$ is not sufficient and can often be improved.
Example 3.25

Let $X_1, \ldots, X_n$ be i.i.d. from the uniform distribution on $(\theta_1, \theta_2)$, $-\infty < \theta_1 < \theta_2 < \infty$.

Note that

$$EX_1 = (\theta_1 + \theta_2)/2 \quad \text{and} \quad EX_1^2 = (\theta_1^2 + \theta_2^2 + \theta_1 \theta_2)/3.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$ and substituting $\theta_1$ in the second equation by $2\hat{\mu}_1 - \theta_2$ (the first equation), we obtain that

$$(2\hat{\mu}_1 - \theta_2)^2 + \theta_2^2 + (2\hat{\mu}_1 - \theta_2)\theta_2 = 3\hat{\mu}_2,$$

which is the same as

$$(\theta_2 - \hat{\mu}_1)^2 = 3(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta_2 > EX_1$, we obtain that

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} + \sqrt{\frac{3(n-1)}{n} S^2}$$

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} - \sqrt{\frac{3(n-1)}{n} S^2}.$$

These estimators are not functions of the sufficient and complete statistic $(X_{(1)}, X_{(n)})$. 
Example 3.26

Let $X_1, ..., X_n$ be i.i.d. from the binomial distribution $Bi(p, k)$ with unknown parameters $k \in \{1, 2, ...\}$ and $p \in (0, 1)$. Since

$$EX_1 = kp$$

and

$$EX_1^2 = kp(1 - p) + k^2 p^2,$$

we obtain the moment estimators

$$\hat{p} = (\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2)/\hat{\mu}_1 = 1 - \frac{n-1}{n} S^2 / \bar{X}$$

and

$$\hat{k} = \hat{\mu}_1^2 / (\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2) = \bar{X} / (1 - \frac{n-1}{n} S^2 / \bar{X}).$$

The estimator $\hat{p}$ is in the range of $(0, 1)$. But $\hat{k}$ may not be an integer. It can be improved by an estimator that is $\hat{k}$ rounded to the nearest positive integer.
Consider the estimation of the central moments
\[ c_j = E(X_1 - \mu_1)^j = \sum_{t=0}^{j} \binom{j}{t} (-\mu_1)^t \mu_{j-t}, \quad j = 2, \ldots, k. \]

the moment estimator of \( c_j \) is
\[ \hat{c}_j = \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \hat{\mu}_{j-t} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j, \quad j = 2, \ldots, k, \]
which are sample central moments, (\( \hat{\mu}_0 = 1 \)).

From the SLLN, \( \hat{c}_j \)'s are strongly consistent.

If \( E|X_1|^{2k} < \infty \), then
\[ \sqrt{n}(\hat{c}_2 - c_2, \ldots, \hat{c}_k - c_k) \rightarrow_d N_{k-1}(0, D) \]

where the \((i,j)\)th element of the \((k-1) \times (k-1)\) matrix \( D \) is
\[ c_{i+j+2} - c_{i+1} c_{j+1} - (i+1)c_i c_{j+2} - (j+1)c_{i+2}c_j + (i+1)(j+1)c_i c_j c_2. \]