11.2 The Horvitz-Thompson Estimator (1952).

The Horvitz-Thompson estimator can be constructed for sampling without replacement as well as sampling with replacement.

How to sample with replacement?

Suppose there are 6 clusters in the population and we want to sample 3 with replacement. Number the clusters 1,2,3,4,5, and 6, and go through the table of random digits ignoring 7,8,9, and 0.

51799153340252

We pick 5, 1, 1 for the sample.

Definition. The Horvitz-Thompson estimator of the population total Y is

$$y'_{hte} = \sum_{i=1}^{\nu} \frac{Y_i}{\pi_i}$$

where Y_i is the total for the ith sampled cluster, π_i is the probability of the ith cluster being selected, and ν is the total number of distinct clusters sampled.

Note that in sampling without replacement, $\nu = m$.

For any single-stage cluster sampling (with or without replacement), it is an unbiased estimator with standard error

$$SE(y'_{hte}) = \sqrt{\sum_{i=1}^{M} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{i=1}^{M} \sum_{j \neq i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right) Y_i Y_j}$$

$$= \sqrt{\sum_{all\ samples} \left(y'_{hte} - \mathbb{E}(y'_{hte})\right)^2 \pi_{ij}} \ \text{(definition)}\,.$$

The following estimator is an unbiased estimator of $\mathbb{V}ar(y'_{hte})$

$$\widehat{\mathbb{V}ar}(y'_{hte}) = \sum_{i=1}^{\nu} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{i=1}^{\nu} \sum_{j \neq i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right) \frac{Y_i Y_j}{\pi_{ij}}$$

The disadvantages of this estimator of the variance are: (1) unstable, (2) can be negative, (3) depends on π_{ij}'s which are difficult to assess.

A more practical estimator is

11.2.2 The Hansen-Hurwitz Estimator (1943)

The Hansen-Hurwitz estimator is

$$y'_{hh} = \frac{1}{m} \sum_{i=1}^{m} \frac{Y_i}{\pi'_i}$$

where π'_i denotes the probability of drawing cluster i.
It is an unbiased estimator of the population total Y for sampling with replacement. Its standard error is

$$SE(y'_{hh}) = \sqrt{\frac{1}{m} \sum_{i=1}^{M} \left(\frac{Y_i}{\pi'_i} - Y \right) \pi'_i}$$

and its estimator is

$$\widehat{SE}(y'_{hh}) = \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left(\frac{Y_i}{\pi'_i} - y'_{hh} \right)^2}.$$

Consider example on pages 338-339 and tables 11.2-11.5.