2021-09-23 DRAFT

Annica de Groote

A Generalized Horvitz-Thompson Estimator

Consider a population consisting of units labeled . Let the th unit be represented by its label and denote the finite population as .

For unit , is the value of the study variable .

We select a probability sample of size from the population . The sampling can be with our without replacement. We are looking for an unbiased estimator of the population total

Let denote the number of times unit is selected in the sample . Let denote the expected value of . For two units and of sizes and respectively, let denote the expected value of the product .

A generalized Horvitz-Thompson estimator of , sometimes called the multiple-count (MC) estimator, is given by

(see for instance Chromy, 2009, or Grafström et al, 2019). It can be shown that is unbiased for . The variance of is given by

and an unbiased estimator of by

## Sampling without replacement

In the case of sampling without replacement, no unit is selected more than once, and can only take the values 0 or 1. We have

where is the probability of including unit in the sample, and is the probability of including both unit and in the sample.

The MC estimator simplifies into the Horvitz-Thompson (HT) estimator

The variance of is given by

and an unbiased estimator of by

In particular, for simple random sampling without replacement,

;

## Sampling with replacement

In the case of sampling with replacement, the sample can include the same unit more than once. The sample size corresponds to the number of draws, not to the number of unique units. We have

where and are the selection probabilities of unit and , respectively. The MC estimator simplifies into the Hansen-Hurwitz (HH) estimator

The variance of is given by

and an unbiased estimator of by

In particular, for simple random sampling with replacement,

for all .

## Multi-stage sampling

Consider a sampling design in three stages, and that sampling is without replacement in all three stages. We then have first and second order inclusion probabilities for each stage:

|  |  |  |
| --- | --- | --- |
|  | Inclusion probabilities | |
| Stage | First order | Second order |
| I |  |  |
| II |  |  |
| III |  |  |

The HT estimator of with respect to all three stages is given by

An unbiased estimator of the variance of is given by

where .

These formulas can potentially be generalized into MC estimation formulas.

## References

Chromy, J. R. (2009). Some Generalizations of the Horvitz-Thompson Estimator. In JSM Proceedings, Survery Research Methods Section. Alexandria, VA: American Statistical Association. Session 347: Memorial for Daniel G. Horvitz: Pioneering Researcher and Visionary Leader, pp 217-227. Retrieved from <http://www.asasrms.org/Proceedings/y2009/Files/302918.pdf>.

Grafström, A., Ekström, M., Jonsson, B G., Esseen, P-A., Ståhl, G. (2019). On combining independent probability samples. Survey Methodology, 45(2): 349-364.