

Bootstrap Variance Estimation using Statistics Canada Survey Data

December 2015

Household Survey Methods Division

Bootstrap Variance Estimation

The Bootstrap method belongs to a family of replicate-based variance estimation techniques. A detailed discussion of replication methods can be found in Wolter (2007), Rust & Rao (1996) or Lohr (1999). Such methods involve the taking of repeated subsamples, or replicates, from the data, re-computing the weighted survey estimate for each replicate, and the full sample, and then computing the variance as a function of the resulting estimates. Other replication methods include Jackknife and Balanced Repeated Replication (BRR).

Several Statistics Canada surveys rely on the bootstrap method to estimate the sampling variance. The method is accessible through a survey-specific set of bootstrap weights that are often provided directly with the survey microdata.

Software Options

In order to properly estimate the variance of a given estimate using replication methods, several different commercially available software packages are available. Examples include SAS, Stata, SUDAAN, WesVar and Bootvar¹. While the bootstrap method is not often directly supported, the BRR is supported and the functionality can be used to estimate the variance for surveys using the bootstrap method. The bootstrap and BRR differ in the way in which the replicates are built but the variance estimation process is the same for both methods. This is a point that some software documentation fails to mention in great detail (Phillips, 2004).

More information on these software can be found on the following websites:

SAS: <http://www.sas.com>

SAS 9.2 is the first version of SAS that offers some replication approaches to variance estimation in its four survey analysis procedures. The BRR option can be used with user-supplied bootstrap weights in order to obtain bootstrap variance estimates (Phillips, 2004).

Stata: <http://www.stata.com>

Stata 12 offers a full suite of design-based variance estimation options with the svy commands – Taylor series, jackknife, BRR, and bootstrap. The bootstrap option can be used with user-specified survey bootstrap weights, such as those provided with many Statistics Canada surveys, in order to obtain bootstrap variance estimates.

¹Bootvar is a bootstrap variance estimation application developed by Statistics Canada. It was originally developed when alternatives, like those described here, were not available. Although it may still be used by some users, it is currently not maintained by Statistics Canada.

SUDAAN: <http://www.rti.org/sudaan>

SUDAAN 10 offers three variance estimation methods – Taylor series, jackknife and BRR. The BRR option can be used with user-specified bootstrap weights in order to obtain bootstrap variance estimates (Phillips, 2004).

WesVar: <http://www.westat.com>

WesVar carries out various analyses of survey data using replication methods for variance estimation. One of the methods offered is BRR with a Fay adjustment, which, as explained in Phillips (2004), can be used to get bootstrap variance estimates if the bootstrap weight variables are provided by the researcher.

Additionally, any software that offers an analytic procedure or command that can produce weighted estimates of the parameters of interest and also has the flexibility of a programming language may be used iteratively to obtain bootstrap variance estimates.

Bootstrap in Practice

A detailed description (along with examples) of how to implement the bootstrap method with the selected software is found in Statistics Canada's Research Data Centre's Information and Technical Bulletin – "Weighted estimation and bootstrap variance estimation for analyzing survey data: How to implement in selected software" (Gagné, 2014).

<http://www.statcan.gc.ca/pub/12-002-x/2014001/article/11901-eng.htm>

With each software option for estimating the bootstrap variance, certain key files, variables and parameters are required. This information differs from one survey to another and therefore, the survey's user guide should be consulted to obtain the specific details. The information required includes:

Files²:

- **The data file** – The file containing all observations and their characteristics collected by the survey
- **The bootstrap weight file** – The file containing the adjusted weights for each of the bootstrap replicates

Variables:

- **Identification variable(s)** – Unique identifiers for every observation on the data file
- **The survey weight** – The weight to be used in calculating the survey estimates – Ex. WTS_M, WTS_S, etc.

² Most software packages require that the 2 files be merged into 1 analytical file. This can be completed by the user via the unique identification variables present in both files.

- **The replicate weights** – The weights that correspond to the survey weight. There should be one weight for each bootstrap replicate
- **Analysis variables** – These are variables to be used by the researcher in their analysis. These may be directly available on the data files or derived for the purposes of the study.

Parameters:

- The replicate weights' names or prefix (Ex. BSW1-BSW1000)
- The number of replicate weights specific to the survey
- A mean bootstrap parameter (Fay adjustment or Fay's Factor) if the survey used the mean bootstrap methodology

References

- Gagné, C. R. (2014, August). Weighted estimation and bootstrap variance estimation for analyzing survey data. *The Research Data Centres Information and Technical Bulletin*.
- Lohr, S. (1999). *Sampling: Design and Analysis*. USA: Duxbury Press.
- Phillips, O. (2004). Using Bootstrap Weights with WesVar and SUDAAN 1(2). *The Research Data Centres Information and Technical Bulletin 1 (2)*, pp. 6-15.
- Rust, K., & Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research 5*, pp. 283-310.
- Wolter, K. (2007). *Introduction to Variance Estimation* (2nd ed.). New York: Springer-Verlag.