



Tips for Working Effectively with RDC Data



Li Wang, May 10, 2023

Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

Canada

Tip 1: Find Data and Information

Data

- RDC data holdings <https://www.statcan.gc.ca/eng/rdc/data>
- Access to some codebooks <http://www62.statcan.ca/webview/>

MAP

- <https://www.statcan.gc.ca/eng/rdc/process>

CRDCN

- Canadian Research Data Centre Network <https://crdcn.ca>



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Tip 2: Key documentation and information

User Guideline

- Level of analysis: National estimates only (CHMS), provincial estimates (CHSCY).
- Changes across data waves: CCHS, CHMS, NPHS, LAD and so on.
- Guideline for analysis (weighting, rounding, variance estimation)

Codebook (variables, values, distributions)

Vetting rules: minimum cell counts, geography level



Tip 3: Strategy for linking large datasets

Strategy 1: make a smaller subset

- start to select target sample, e.g., age>40 or year>2018
- keep the variables with interest + weight variables
- use the cohort with interest as the base to link with other data source
e.g., CCHS-IMDB, only keep the immigration sample
- use 1:1 merge to drop observations that were not linkable



Tip 3: Strategy for linking large datasets

Strategy 2: use macro or loop to merge panel data (e.g., IMDB_T1FF)

- use SAS – Data format, process speed
- select variables, rename and append
- use macro or loop to make the process simple
- ❖ Ask for help from Analysts

Tip 4: Weights

- ❑ Weighting is used to make samples **match the population**. Weighting is necessary when each observation has a different probability of selection in a sample.

- ❑ Understanding different weights
 1. **pweight: Sampling weight.**
 - Each observation is treated as a randomly selected sample from the group which has the size of weight.
 - Most weights on RDC data are pweight.

 2. **aweight: Analytic weight.**
 - Each observation is treated as the mean of a group which has the size of weight.
 - Researcher needs to calculate the aweight: $AW_i = pweight_i * (\text{sample size } n / \text{Weighted } N)$

 3. **fweight: Frequency weight (= weight in SPSS).**
 - By weighting, each observation is duplicated by the size of weight.
 - Integer of the pweight: $\text{round}(pweight)$



Tip 4: Weights – continued

- Weighting must be applied. Otherwise, your results are biased in most cases.
 - In Stata, the estimated coefficients are identical regardless of which weighting options are applied. All weighting methods will yield the same coefficients estimated.
 - But standard errors vary by weighting options.
 - Usually, SE errors with pweight > SE with aweight > SE with fweight
- Use pweight for modeling
- Use aweight for descriptive statistics.
- Use fweight for population projection.
 - e.g., if vetting requires rounding, use fweight to get the numerator and dominators.
- Don't use iweight (important weight).

Tip 5: Household vs. Individual Level Weights

- ❑ Many datasets have both a household and an individual level weight (CCHS, OCHS)
 - Interview surveys are often sampled and conducted at the household level.
 - One respondent, usually at random, is selected to be interviewed.
 - Considering the differential selection of individuals in households
 - For household with only one adult the sampling fraction is 1/1
 - For household with 3 adults the fraction is 1/3
- ❑ Use household weight when you want to generalize to characteristics of households (e.g., poverty, parent marital status)
- ❑ Use individual (person) weight when generalizing to a population of individuals

Tip 7: Bootstrap weights

- ❑ Encouraging **estimating variances** using the bootstrap weights.
 - To account for the complex survey design
 - Unbiased SEs
- ❑ Appropriate survey weight variable and corresponding bootstrap weight variables
 - HH level vs. Individual level (CCHS, OCHS)
- ❑ Whether bootstrap weight variables are mean bootstraps, and, if so, the number of replicates in each mean bootstrap weight (needed for a bootstrap weight adjustment)
 - e.g. , GSS, OCHS
 - check if there is 0 in bootstrap data or description in the user guideline.
 - <https://www150.statcan.gc.ca/n1/pub/12-002-x/2014001/article/11901-eng.htm>

Thank You!

Further Questions:

Li Wang, rdc4@mcmaster.ca