Weighting the Alberta EDI Community Data

Vijaya Krishnan, Ph.D.

2 October 2011

What is Weighting?

Weighting is a way of emphasizing some aspects of a dataset (e.g., non-response) to alter the end result so that in our population we can expect as many cases of this aspect as there are in our study sample. This is crucial because a representative subgroup of a population is important if we want to make any inference.

The EDI is a population-based survey. However, despite our best efforts, we sometimes end up having children of a certain characteristic, such as those who speak English/French as a first language more than those who speak languages other than English or French, differing from the way it is distributed in the population. The sex distribution of the study population can also be an important factor. We have found that females are likely to have higher EDI scores than males. Children in a sub-community, say A, may consist of 30 percent girls, when girls make up of 50 percent for the community to which A belongs. This introduces bias into any estimate we may obtain from our (sub) community analysis because statistical procedures will give greater weight to those we over-survey. We can correct for such biases mathematically with the weighting process.

Why Weight?

Post-stratification Weights

We weight the EDI community data for three distinct reasons using post-stratification weights. Before we discuss these reasons, let us briefly discuss what is meant by post-stratification weight.

A post-stratification weight got its name from the fact that we can only compute it after we have collected the data. The stratification comes from the idea that we use various strata or characteristics (e.g., age, sex, etc.) to adjust our data to conform more closely to the population. Ideally, if we are studying children in the age group, say 4-7 years of age in a community, we would need all children in the same age-group for the province in order to calculate a post-stratification weight variable age. This being almost impossible to attain, we resort to the use of community data from the same survey to compute post-stratification weights. We could compare our sub-community data to the community data, to make sure that the distribution of characteristics, such as sex, E/FSL status, and non-response rate (if available) are similar to that of the community data. If the distributions are fairly close, there is no need to worry about weighting at all. If not, weighting needs some consideration.

The Reasons for Weighting

First, the sub-community level data may not be representative of the community without the weights. The sub-communities are drawn for practical reasons using Dissemination Area boundaries. This may result in over-or under count of children of a particular characteristic resulting in more or less number of observations. Small n's are always a cause for concern in sub-community analysis because it is difficult to carry out any meaningful analysis. Weights that correct such bias are often termed as design weights. In general, weighing corrects for skewness in representing a population.

Second, let us assume that we have a sub-community that is representative of the community. What if, a large proportion of questionnaires in the sub-community include *no active consent* ones or proportionately fewer valid cases? To account for no-consent to the survey, adjust for the parents who did not respond to the survey (non-

response adjustment). If groups differ in their response rates (RR), then a non-response adjustment is necessary. In other words, when the RRs are unequal, include a weighing factor proportional to the inverse of the RR. In simple terms, mathematically, we can correct biases of this kind with what are called non-response weights or a proxy for it.

Finally, weighting can be applied if we want our results to reflect as results that would have attained if a sub-community was the same as a community to which the sub-community is a part of.

Possible Weight Variables for the EDI Community Data

We are constrained in the use of weighting variables for the community data. One of the reasons is, variables, such as age or days absent have distributions with relatively little or no variation, compared to sex or ESL status. Based on some preliminary analyses, we are convinced that it is highly important to look at three variables: sex, E/FSL status, and valid EDI (after all filters are applied).

Procedure to Compute Post-stratification Weights

1. Run frequencies of variables (e.g., sex, E/FSL status) for the unweighted data
2. Compute weights for each variable as in Table 1, making use of the total count
3. Combine the weighted variables (e.g., *sexwt*, *efslwt*) into one weight by multiplying them since only one weight can be used in the analysis
4. Re-base the weight variable *wt* to ensure that the size is maintained at the community level (*rebasedwt*).
5. Weight the cases by *rebasedwt* before conducting the analysis.

The example below gives an illustration of how to compute post-stratification weights for sex, E/FSL, and non-response rates. The non-response rate is a proxy and it represents the variability in questionnaires analyzed across sub-communities. That is, it is not based on parental consent alone; all those cases in the file minus those that were taken out due to reasons, such as missing a developmental area, special needs, less than 1 month in school, etc.

NOTE: For the example below, the weighting is done at the sub-community level using three variables: sex (Male/Female); E/FSL (E/F, first language/E/F second language), and response rate (as a proxy based on questionnaires analyzed vs. questionnaires received). A diagnostic check on the importance of the variables in child developmental outcomes is conducted prior to deciding the weight variables. Although, age is another important variable, the discussion below excludes this variable because the purpose here is to introduce the procedure.

ECMap Subcommunity Number (ECMapSubN)	Sex (egender; Female, code=1; Male, code=2)				E/FSL (efsl; code=0; and 1 and 2)				Response Rate (RR)			Non-Response Rate
[Col 1]	Female [Col 2]		Male [Col 3]		E/F first language [Col 4]		E/F is a second language [Col 5]		Qs analyzed* [Col 6]	Qs received [Col 7]	RR Col 6/Col 7 [Col 8]	(1/RR) 1/Col 8 [Col 9]
1	9	31.03	20	68.97	20	95.23	1	4.77	18	29	0.621	1.610
2	70	47.30	78	52.70	128	87.67	18	12.33	123	148	0.832	1.202
3	118	47.77	129	52.23	214	91.06	21	9.94	202	247	0.818	1.222
4	131	48.16	141	51.84	234	88.30	31	11.70	221	272	0.813	1.230
5	157	49.37	161	50.63	278	88.82	35	11.18	247	318	0.777	1.287
6	82	48.52	87	51.48	164	97.04	5	2.96	147	169	0.870	1.149
7	97	44.50	121	55.50	174	85.71	29	14.29	164	218	0.752	1.330
8	107	52.45	97	47.55	190	93.60	13	6.40	182	204	0.892	1.121
Edmonton West	771	48.04	834	51.96	1402	90.16	153	9.84	1304	1605	0.812	1.232

*filters are in place

Creating a post-stratification weight in SPSS of the sex variable

The syntax for constructing a weight variable sex is:

```
If (egender=1 & ECMapSubN=1)sexwt=31.03.
If (egender=2 & ECMapSubN=1)sexwt=68.97
If (egender=1 & ECMapSubN=2)sexwt=47.30.
If (egender=2 & ECMapSubN=2)sexwt=52.70.
If (egender=1 & ECMapSubN=3)sexwt=47.77.
If (egender=2 & ECMapSubN=3)sexwt=52.23.
If (egender=1 & ECMapSubN=4)sexwt=48.16.
If (egender=2 & ECMapSubN=4)sexwt=51.84.
If (egender=1 & ECMapSubN=5)sexwt=49.37.
If (egender=2 & ECMapSubN=5)sexwt=50.63.
If (egender=1 & ECMapSubN=6)sexwt=48.52.
If (egender=2 & ECMapSubN=6)sexwt=51.48.
If (egender=1 & ECMapSubN=7)sexwt=44.50.
If (egender=2 & ECMapSubN=7)sexwt=55.50.
If (egender=1 & ECMapSubN=8)sexwt=52.45.
If (egender=2 & ECMapSubN=8)sexwt=47.55.
Execute.
```

The syntax for constructing an E/FSL variable is:

```
If (efsl=0 & ECMapSubN=1)efslwt=95.23.
If (efsl=1I efsl=2 & ECMapSubN=1)efslwt=4.77.
If (efsl=0 & ECMapSubN=2)efslwt=87.67.
If (efsl=1I efsl=2 & ECMapSubN=2)efslwt=12.33.
If (efsl=0 & ECMapSubN=3)efslwt=91.06.
If (efsl=1I efsl=2 & ECMapSubN=3)efslwt=9.94.
If (efsl=0 & ECMapSubN=4)efslwt=88.30.
If (efsl=1I efsl=2 & ECMapSubN=4)efslwt=11.70.
If (efsl=0 & ECMapSubN=5)efslwt=88.82.
If (efsl=1I efsl=2 & ECMapSubN=5)efslwt=11.18.
If (efsl=0 & ECMapSubN=6)efslwt=97.04.
If (efsl=1I efsl=2 & ECMapSubN=6)efslwt=2.96.
If (efsl=0 & ECMapSubN=7)efslwt=85.71.
If (efsl=1I efsl=2 & ECMapSubN=7)efslwt=14.29.
If (efsl=0 & ECMapSubN=8)efslwt=93.60.
If (efsl=1I efsl=2 & ECMapSubN=8)efslwt=6.40.
Execute.
```

The syntax for constructing a non-response rate weight is:

```
If (ECMapSubN=1)nonresponsewt=1.610.
If (ECMapSubN=2)nonresponsewt=1.202.
If (ECMapSubN=3)nonresponsewt=1.222.
If (ECMapSubN=4)nonresponsewt=1.230.
If (ECMapSubN=5)nonresponsewt=1.287.
If (ECMapSubN=6)nonresponsewt=1.149.
If (ECMapSubN=7)nonresponsewt=1.330.
If (ECMapSubN=8)nonresponsewt=1.121.
Execute.
```

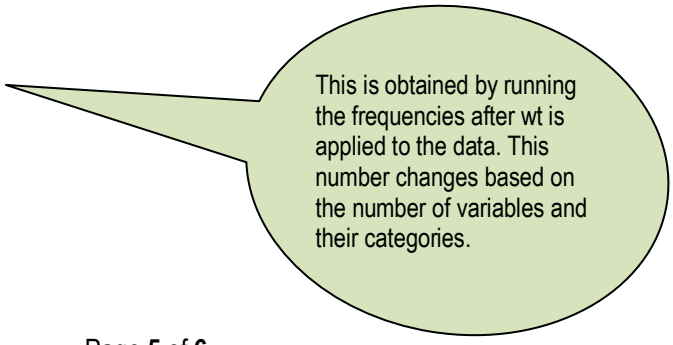
Compute wt=sexwt*efslwt*nonresponsewt.

Normalization of weights is another important step in weighting. The process sets the weights so the N in the weighted data equals the N in the unweighted data. Re-basing the weight variable returns the overall N for the community (N=1304 for Edmonton West). The weighted sample using the rebased weight allows us to work with the smaller N. The process involves running frequencies applying the weight variable before re-basing it.

Rebasedwt= wt*(the number of cases in the raw data file/the number of cases when the weight variable *wt* is applied)

The syntax for re-basing the weight:

```
Compute rebasedwt= wt* (1304 / 6619749)
Execute.
```



References

Debell, M. & Krosnick, J. A. 2009. Computing weights for American National Election Study Survey Data, http://www.electionstidies.org.

Johnson, D. R. Using weights in the analysis of survey data, Population Research Institute, Pennstate.

Norman, P. and Wathan, J. 2010. Using SPSS for Windows: Exploring the Health Survey for England 2002, Teaching Dataset in SPSS v15/16, Economic and Social Data Service.

PAHO/WHO. Weighting your data. www.paho.org/english/ad/dpc/nc/steps-tng-10-**data**-**weighting**.ppt

University of Regina, Weight variable in the General Social Survey (GSS): how should they be used? Retrieved on: Oct 1, 2011, *http://uregina.ca/datalibrary/weight.html*