Methodological Note #007

Survey Weights in AmericasBarometer Data

Oscar Castorena
Vanderbilt University

January 26, 2021

Key Findings:

- I describe the survey weights included in AmericasBarometer data, along with explanations of the types of adjustments those weights make to the raw survey data.
- I provide recommendations about how to properly use the different weights in the data given the different analyses a researcher might conduct.
- An illustrative analysis is used to highlight the implications of omitting weights in survey analysis, which can affect both point estimates and standard errors.

USAID FROM THE AMERICAN PEOPLE
VANDERBILT
LAPOP
AmericasBarometer
Barómetro de las Américas

Weights are a common feature of survey data. They allow researchers to make inferences about populations of interest that may otherwise be flawed if using simple analyses of raw survey data. Since such inferences from samples are the aims of survey research, then understanding the purpose and appropriate implementation of survey weights, as well as other features of complex survey designs, is an important step in the analysis of public opinion data. However, a recent meta-analysis finds that analytic errors (i.e. analyses that ignore complex design features such as weights, stratification, and cluster sampling) are extremely prevalent in secondary analyses of survey data.[1] Moreover, as West, Sakshaug, and Kim (2017) argue, these analytical errors undermine the considerable resources invested in minimizing many sources of error in the collection of survey data as summarized by the total survey error paradigm.[2] This *Methodological Note* is meant to help prevent such analytical errors by providing an overview of the weights included in AmericasBarometer data. This overview classifies the different weights and discusses the adjustments they make to the raw survey data. A table summarizing the use of weights for all countries and waves of the AmericasBarometer is included in the supplementary appendix material of this note. Moreover, the note provides an illustrative example of the implications for analyses that ignore the complex design of AmericasBarometer surveys.

Description of AmericasBarometer Weights

In this section, I describe the different types of weights used in AmericasBarometer data in terms of the adjustment they aim to make. I use the surveys from Ecuador as illustrative examples of different weights/adjustments in the AmericasBarometer data. The supplementary appendix material to this note includes Stata code for these examples. Ecuador is a useful case, since throughout the time series, weights have been used for different purposes. The Ecuador time series also includes self-weighted samples.

The weights included in AmericasBarometer survey data generally aim to make one of three forms of adjustments and, depending on the analysis, a combination of the three. The first type of weight is a *post-stratification adjustment*. In this case, the final sample of completed interviews deviates from the sample design on key demographic benchmarks. Most AmericasBarometer surveys use frequency matching for the selection of respondents at the household level, which results in these samples typically not requiring a post-stratification adjustment.[3] A few samples, however, use random selection at the level of each household. As a result, the overall sample may not align with known population margins. One common example is deviations in the gender composition of a sample since male respondents are less likely to be home during fieldwork hours.[4] These deviations necessitate the use of weights so that weighted margins on key demographic variables (typically sex, age group, and/or urban/rural geographic location) agree with known population proportions. For the AmericasBarometer, the post-stratification is implemented at the level of the primary strata (the estratopri variable in the dataset) such that weighted margins are in line with known population proportions at the primary strata level. This is important because AmericasBarometer surveys are designed to be representative at the level of the primary strata.

The AmericasBarometer surveys for Ecuador from the 2004 and 2006 waves include weights that make post-stratification adjustments. The data collection for these surveys used random selection at the level of each household and as a result have gender imbalances in the final sample. Table 1 summarizes the percentages of each gender category for the raw samples of the Ecuador surveys (2004-2012). In the first two waves, there is an overrepresentation of female respondents. The surveys conducted after the 2006 wave used frequency matching at the household level and thus do not have such imbalances in the final sample. Table 2 summarizes the gender distributions using the weights included in the data sets. Notice that the post-stratification adjustment of the weights brings the gender distributions of the first two waves in line with population benchmarks. Any analysis of the marginal distributions of variables from the 2004 and 2006 Ecuador surveys that ignores weights may result in biased estimates if the variable of interest is correlated with gender.

Table 1: Gender Distributions of Ecuador Surveys Raw Data

	2004	2006	2008	2010	2012
Male	40.47	42.53	50.00	49.22	49.93
Female	59.53	57.47	50.00	50.78	50.07

Note: 2004 N = 3,000; 2006 N = 2,925; 2008 N = 3,000; 2010 N = 2,999; 2012 N = 1,500.

Table 2: Gender Distributions of Ecuador Surveys Using Weights

	2004	2006	2008	2010	2012
Male	49.00	49.00	49.00	49.00	49.93
Female	51.00	51.00	51.00	51.00	50.07

Note: 2004 N = 3,000; 2006 N = 2,925; 2008 N = 3,000; 2010 N = 2,999; 2012 N = 1,500.

A second type of weight common in AmericasBarometer survey data is one meant to adjust for *oversamples* that were part of the survey design. Unlike the post-stratification adjustment described above, this type of weight is a planned feature of the final dataset prior to fieldwork. LAPOP Lab will occasionally include oversamples of subpopulations of interest in their surveys. These may be necessary if the lab or local partners are interested in studying the attitudes of subpopulations whose representation in a typical nationally representative sample may be too small for generating precise estimates of opinions or relationships between variables. One example of this kind of oversample is the 2010 Chile survey, which included an oversample of areas affected by a recent earthquake. Since those earthquake-affected areas are then overrepresented, a weight is necessary for any analysis that seeks to make inferences about Chile nationally. Specifically, that weight will down-weight responses from earthquake-affected areas (the oversample) and place greater weight on other areas.

The AmericasBarometer surveys for Ecuador from the 2004 to 2010 waves include weights that make adjustments for oversamples in the survey design. In this case, the east region, which includes the Amazon, was oversampled. Since this area is sparsely populated, a nationally representative sample will typically include too few respondents to be able to make precise statistical inferences about that subpopulation. In the AmericasBarometer time series, the east region was oversampled from 2004 to 2010 in order to permit such inferences. Table 3 summarizes the unweighted (or raw) regional distributions of the Ecuador surveys (2004-2012). Contrast these distributions with the weighted percentages in Table 4, which brings the distribution in line with population benchmarks. Note that the estimates for 2012 (which included no oversample and is a self-weighted sample) are identical for weighted and unweighted distributions. Any analysis of the marginal distributions of variables from 2004 to 2010 Ecuador surveys that ignores weights may result in biased estimates if the variable of interest is correlated with geographic region.

Table 3: Regional Distributions of Ecuador Surveys Raw Data

	2004	2006	2008	2010	2012
Coast	44.23	44.21	44.23	43.88	50.33
Highlands	39.83	39.45	39.83	39.71	45.00
East	15.93	16.34	15.93	16.41	4.67

Note: 2004 N = 3,000; 2006 N = 2,925; 2008 N = 3,000; 2010 N = 2,999; 2012 N = 1,500.

Table 4: Regional Distributions of Ecuador Surveys Using Weights

	2004	2006	2008	2010	2012
Coast	50.91	50.58	49.58	49.20	50.33
Highlands	45.08	45.39	46.31	46.66	45.00
East	4.00	4.03	4.11	4.14	4.67

Note: 2004 N = 3,000; 2006 N = 2,925; 2008 N = 3,000; 2010 N = 2,999; 2012 N = 1,500.

Either one or a combination of the two previous adjustments are what make up the samples classified as weighted in LAPOP's technical information for each survey. The weighting variable to be used for single country-year analyses is the "wt" variable in each dataset. The majority of AmericasBarometer surveys are self-weighted. This means they do not require any of the previously discussed adjustments and all observations in those data sets have a "wt" variable with a value of one (all responses have equal weight).[5]

The final type of weight included in AmericasBarometer data sets is meant for analyses of *multiple surveys* (either multiple countries or multiple waves). The weighting variable to be used for these kinds of analyses is the "weight1500" variable in each dataset. Since some surveys (due to oversamples) may have considerably more than 1,500 respondents, they may be more influential than others in making inferences about region-wide opinion. The "weight1500" adjustment treats each survey as though it had 1,500 respondents when conducting pooled analyses.[6] It should be noted that this weight adjusts region-wide or pooled analyses such that each country has equal weight and does not weigh each country proportional to their population. For example, when using weight1500 to conduct pooled analyses, the Uruguay sample is as influential as the Brazil sample.[7]

It is recommended that whenever a researcher is working with a merged dataset of any kind (i.e. a dataset that contains more than one study whether that is multiple countries or multiple waves), they use "weight1500" as the weighting variable. This includes instances where the researcher is primarily using the merged dataset to analyze a single study. In such cases, "weight1500" will not affect any inferences made from a single study analysis. This is because "weight1500" is applied uniformly for all respondents in a given study, multiplying the "wt" variable by a constant value. Therefore when analyzing a single study, using "weight1500" will be functionally equivalent to using "wt".[8]

Implications of (Non-)use of Weights in Survey Analysis

The previous section provided an overview of the nature and purpose of weights in AmericasBarometer survey data. Continuing with the example of Ecuador, this section will illustrate the implications for analyses that do not account for weights in the data and, more broadly, the complex survey design of an AmericasBarometer study. Since the AmericasBarometer includes a core set of questions in each wave, which goes back to 2004 for some countries, the data is especially useful for researchers interested in public opinion patterns over time. For this reason, I conduct two analyses of a survey item over time, one taking into account survey weights and the complex design of the survey[9] and one that treats the data as if it were a simple random sample.

I focus on a survey question about neighborhood insecurity (**AOJ11**) included since 2004. The question asks, "Speaking of the neighborhood where you live and thinking of the possibility of being assaulted or robbed, do you feel very safe, somewhat safe, somewhat **un**safe or very **un**safe?" Figure 1 plots the percent of respondents who answer either very unsafe or somewhat unsafe over the different waves of the survey (here I focus on the period between 2004 and 2012). The circle symbols represent estimates from a design-based analysis (which includes weights) and the square symbols represent estimates from an unweighted analysis that ignores the complex survey design.

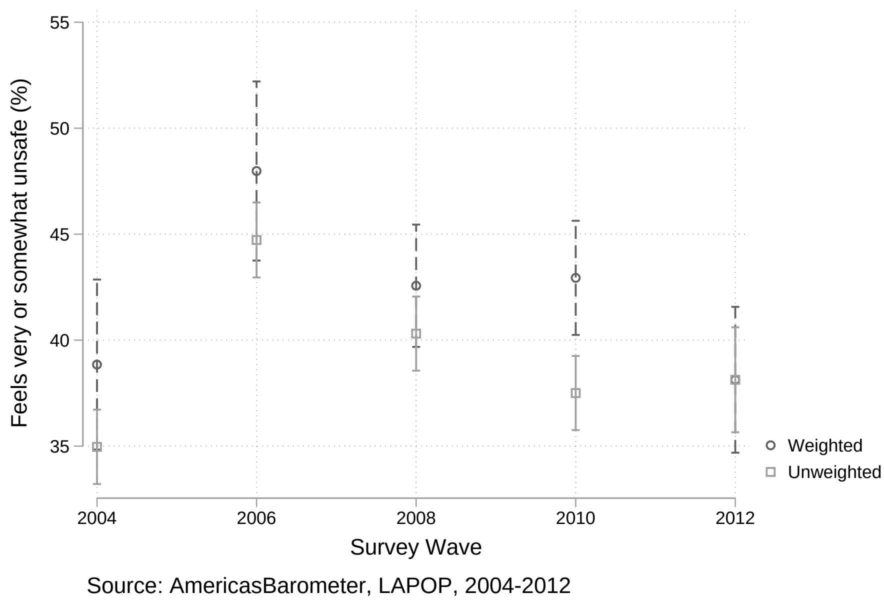

Figure 1: Perception of Neighborhood Insecurity, Ecuador 2004-2012

The first pattern to note about the figure is that from 2004 to 2010 the unweighted estimate is lower than the weighted estimate. As discussed earlier, the Ecuador surveys from 2004 to 2010 included an oversample of the eastern region of the country. Since this area of the country is overrepresented in these waves of the survey, any analyses of variables with significant regional differences may have biased estimates as a result. This appears to drive the downward bias in estimates of neighborhood insecurity in Figure 1. The eastern region, which includes the Amazon and is generally more rural than the rest of the country, has considerably lower levels of neighborhood insecurity. For the period between 2004 and 2012, 30.5% of respondents in the eastern region reported feeling unsafe, compared to 42.6% for the rest of the country. The overrepresentation of respondents in the region results in downwardly biased estimates in unweighted analyses.

A second thing to notice about the figure is the 95% confidence intervals around each estimate (weighted and unweighted). The difference between these quantities is most apparent in the estimates for the 2012 wave. In this case, the survey is self-weighted and thus the weighted and unweighted analyses produce the same estimate. However, it is also apparent that the two analyses produce different confidence intervals. The reason for this is that the unweighted analysis treats the data as if it were a simple random sample whereas the weighted estimates are from an analysis that accounts for the complex design of the AmericasBarometer, which includes stratification, clustered samples, and in some cases weighting.[10] This design-based analysis accounts for the effects of this sample design in estimating variances and generating confidence intervals. In this case, the confidence interval from a design-based analysis is larger than from an unweighted analysis.[11] These differences in variances and confidence intervals could significantly affect inferences researchers are trying to make from the survey data.

Conclusion

This note provided an overview of the different types of weights included in AmericasBarometer survey data, including weights meant as post-stratification adjustments, weights meant to adjust for oversamples in the survey design, and weights meant to adjust for cross-country and/or cross-time analyses. These weights are summarized in the supplementary appendix material for all countries and waves of the AmericasBarometer. Using the example of Ecuador surveys, the note demonstrates the implications for analyses that ignore weights as well as the overall complex design of the survey. The example of neighborhood insecurity in Ecuador over the period 2004 to 2012 demonstrates that unweighted analyses can result in biased estimates. Moreover, the differences in confidence intervals for the estimates from the 2012 wave demonstrate the importance of design-based analyses even when working with self-weighted samples. This point is especially important when considering that the great majority of surveys in the AmericasBarometer series are self-weighted. Even when working with such self-weighted samples, researchers should rely on design-based analysis of the data.

Table 5: Summary of Weights in AmericasBarometer Data

Country	2004	2006	2008	2010	2012	2014	2016	2018
Mexico	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Guatemala	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
El Salvador	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Honduras	self-weighted	self-weighted	self-weighted	self-weighted	weighted^{+}	self-weighted	self-weighted	self-weighted
Nicaragua	self-weighted	self-weighted	self-weighted	self-weighted	weighted^{+}	self-weighted	self-weighted	self-weighted
Costa Rica	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Panama	self-weighted	self-weighted	self-weighted	self-weighted	weighted^{+}	self-weighted	self-weighted	self-weighted
Colombia	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Ecuador	weighted^{*+}	weighted^{*+}	weighted^{+}	weighted^{+}	self-weighted	self-weighted	self-weighted	self-weighted
Bolivia	weighted^{+}	weighted^{+}	weighted^{+}	weighted^{+}	weighted^{+}	weighted^{+}	self-weighted	self-weighted
Peru	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	weighted^{+}	self-weighted
Paraguay	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Chile	–	self-weighted	self-weighted	weighted^{+}	weighted*	weighted*	self-weighted	self-weighted
Uruguay	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Brazil	–	self-weighted	self-weighted	weighted^{+}	weighted^{+}	weighted^{+}	weighted^{+}	weighted^{+}
Venezuela	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	–
Argentina	–	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Dominican Republic	weighted*	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Haiti	–	self-weighted	self-weighted	weighted^{+}	weighted^{+}	self-weighted	weighted^{+}	–
Jamaica	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted
Guyana	–	self-weighted	weighted^{+}	self-weighted	self-weighted	self-weighted	weighted^{+}	–
Trinidad and Tobago	–	–	–	weighted^{+}	weighted^{+}	weighted^{+}	self-weighted	–
Belize	–	–	self-weighted	self-weighted	self-weighted	self-weighted	self-weighted	–
United States	–	self-weighted	weighted*	weighted*	weighted*	weighted*	weighted*	weighted*
Canada	–	self-weighted	weighted*	weighted*	weighted*	weighted*	weighted*	weighted*
Suriname	–	–	–	weighted^{+}	self-weighted	weighted^{+}	–	–
Bahamas	–	–	–	–	–	weighted^{+}	–	–
Barbados	–	–	–	–	–	weighted^{+}	–	–
Grenada	–	–	–	–	–	–	self-weighted	–
Saint Lucia	–	–	–	–	–	–	self-weighted	–
Dominica	–	–	–	–	–	–	self-weighted	–
Antigua and Barbuda	–	–	–	–	–	–	self-weighted	–
St. Vincent and the Grenadines	–	–	–	–	–	–	self-weighted	–
St. Kitts and Nevis	–	–	–	–	–	–	self-weighted	–

* post-stratification adjustment $^{+}$ weight to adjust for oversample in design

Appendix: Stata Code Examples for Using Weights in AmericasBarometer Data

Syntax for Results in Methodological Note

Stata includes a suite of commands and features that facilitate working with complex survey data. In order to make use of these features one must first declare the survey design for the dataset. This is done using the `svyset` command.[12] When using a merged dataset (i.e. cross-country and/or cross-time datasets), as in the analyses for this note, the specific syntax and output from Stata are the following:

```
. svyset upm [pw=weight1500], strata(strata)
pweight: weight1500
VCE: linearized
Single unit: missing
Strata 1: strata
SU 1: upm
FPC 1: <zero>
```

While the reader should consult the Stata documentation for details on the `svyset` command, there are a few aspects of the syntax above that are worth noting. First, "upm" is declared as the variable in the dataset that identifies the primary sampling units in the AmericasBarometer data. Remember that, since the AmericasBarometer (2004-2019) is a face-to-face survey, interviews are collected in geographic clusters. This part of the `svyset` command identifies those clusters. Second, as recommended in this note for merged datasets, "weight1500" is set as the sampling weight (`[pw=weight1500]`). Finally, the `strata()` option is used to specify the primary strata in the sample design, which in this case is identified with a variable called "strata" in the AmericasBarometer dataset. As discussed in the note, these three survey design features (weight, clusters, and strata) are consequential for the calculation of both point estimates and standard errors.

Datasets downloaded from LAPOP should come with the survey design settings already saved. One can use the `svydescribe` command to verify that the saved design features (such as the weighting variable and primary strata) are appropriate for the data and analysis.

Now that we have declared the structure of the survey data in Stata, we can begin using `svy`, the survey prefix command, to conduct analyses that take into account the survey design. However, we first show the syntax for the naïve analysis in Table 1 that tabulates the unweighted distribution of gender across the waves of the Ecuador surveys. Note that the syntax below restricts the analysis to observations from Ecuador (which is identified with a value of 9 in the "pais" variable of the dataset) and from the 2012 wave and earlier (this is done using the `if` statement following the tabulate command). These restrictions are necessary in the syntax because we are working with a merged dataset including all countries and waves.

```
.tab q1 wave if pais == 9 & wave <= 2012, col nofreq
```

As discussed in the main text, the above syntax treats the data as though it were a simple random sample rather than a multistage survey with weights, clustering, and stratified sampling. The syntax below produces the results of Table 2, which is a design-based analysis of the distribution of gender across waves of the survey in Ecuador. Note that the only differences with the previous syntax is the addition of the `svy` prefix before the command (which executes a design-based analysis) and the inclusion of the `percent` option (which provides percentages in the output table rather than default proportions when using `svy: tab`). The `svy` prefix greatly simplifies conducting design-based analyses of survey data in Stata.

```
. svy q1 wave if pais == 9 & wave <= 2012, col percent
```

The syntax below reproduces the results for Table 3 in the note, which are unweighted distributions of the regional strata in the Ecuador surveys across the different waves.


```
. tab estratopri wave if pais == 9 & wave <= 2012, col nofreq
```

The correct (i.e. weighted) distributions of the three regions from Table 4 are produced with the syntax below, which uses the `svy` prefix.

```
. svy: tab estratopri wave if pais == 9 & wave <= 2012, col
```

In order to reproduce the results of Figure 1 in the note, one must first recode the "aoj11" variable in the AmericasBarometer dataset to a binary variable as described in the main text. This is done using the syntax below, which generates a new variable called "aoj11_recode".

```
. recode aoj11 (1 2 =0) (3 4=100), gen(aoj11_recode)
```

The unweighted estimates in Figure 1 are generated using the `mean` command. The syntax below ignores the need for survey weights as well as the structure of the dataset which includes clustering and stratification. The output of the command is also provided.

```
. mean aoj11_recode if pais == 9 & wave <= 2012, over(wave)

Mean estimation                    Number of obs = 13,324

_subpop_1: wave = 2004
_subpop_2: wave = 2006
_subpop_3: wave = 2008
_subpop_4: wave = 2010
_subpop_5: wave = 2012

---------------------------------------------------------------

Over          |    Mean      Std. Err.    [95% Conf. Interval]

---------------------------------------------------------------

aoj11_recode |
_subpop_1    |  34.96456  .8761865    33.24711    36.68201
_subpop_2    |  44.7269   .9217111    42.92021    46.53358
```

```
_subpop_3     |   40.30852   .898408     38.54751     42.06953
_subpop_4     |   37.50419   .8868644    35.76581     39.24257
_subpop_5     |   38.13046   1.259984    35.66072     40.60021

----------------------------------------------------------------
```

In order to execute a design-based (i.e. weighted) analysis of cross-time trends in neighborhood insecurity, one simply needs to add the `svy` prefix as in the syntax and output below.

```
. svy: mean aoj11_recode if pais == 9 & wave <= 2012, over(wave)
(running mean on estimation sample)

Survey: Mean estimation
Number of strata = 15                    Number of obs = 13,286
Number of PSUs = 449                Population size = 7,422.2153
Design df = 434


_subpop_1: wave = 2004
_subpop_2: wave = 2006
_subpop_3: wave = 2008
_subpop_4: wave = 2010
_subpop_5: wave = 2012


----------------------------------------------------------------

                            Linearized

Over          |    Mean      Std. Err.    [95% Conf. Interval]

----------------------------------------------------------------

aoj11_recode |
_subpop_1     |   38.84717   2.040934    34.83582     42.85851
```

```
_subpop_2    |   47.98135   2.150574    43.75451    52.20819
_subpop_3    |   42.56732   1.468479    39.6811     45.45354
_subpop_4    |   42.93966   1.370067    40.24687    45.63245
_subpop_5    |   38.13046   1.75099     34.68899    41.57194

----------------------------------------------------------------
```

Consequences of Using the Incorrect Weight

In the main text of this methodological note, it was recommended that researchers always use "weight1500" as the weighting variable when working with merged AmericasBarometer datasets. The following example illustrates the consequences of using the wrong weight variable (in this case "wt" rather than "weight1500"). To do, we make use of a figure referenced in the note.

In the main text's discussion of Figure 1, it was noted that in the period between 2004 and 2012, 30.5% of respondents in the eastern region reported feeling unsafe compared to 42.6% in the rest of the country. These percentages were calculated through a design-based analysis that used the `svy` prefix in Stata. Before running the analysis, an indicator variable for the east region is generated with the syntax below.

```
. gen east = 1 if estratopri == 903 & pais == 9
. replace east = 0 if estratopri != 903 & pais == 9
```

The following syntax and output produces the 30.5% and 42.6% figures reported in the main text. Note that since the weighting variable here is "weight1500", each wave of the Ecuador surveys has equal weight in the calculation of the average for the period. Moreover, this is especially necessary since the Ecuador surveys prior to 2012 had large oversamples.

```
. svy: mean aoj11_recode if pais == 9 & wave <= 2012, over(east)
(running mean on estimation sample)
```

```
Survey: Mean estimation
Number of strata = 15                    Number of obs = 13,286
Number of PSUs = 449                 Population size = 7,422.2153
Design df = 434


0: east = 0
1: east = 1


----------------------------------------------------------------

                           Linearized

Over          |    Mean      Std. Err.    [95% Conf. Interval]

----------------------------------------------------------------

aoj11_recode |
0            |  42.59942   .8310237    40.96609    44.23275
1            |  30.46638   2.507745    25.53754    35.39522

----------------------------------------------------------------
```

To illustrate how failing to use the correct weighting variable for analysis of merged data can affect the results, we use the `svyset` command to set "wt" as the weighting variable in the dataset. The syntax and output are below.

```
. svyset upm [pw=wt], strata(strata)
pweight: wt
VCE: linearized
Single unit: missing
Strata 1: strata
SU 1: upm
FPC 1: <zero>
```


Now we re-run the design-based analysis but with the incorrect weighting variable and get the output below. Rather than the 30.5% as in the analysis above, this analysis reports 31.5% of respondents from the eastern region reported feeling unsafe in their neighborhood for the period 2004-2012. The difference here results from the fact that the earlier waves with more respondents have greater influence in calculating the average for the period compared to the analysis using "weight1500", which weights each wave equally.

```
. svy: mean aoj11_recode if pais == 9 & wave <= 2012, over(east)
(running mean on estimation sample)

Survey: Mean estimation
Number of strata = 15                       Number of obs = 13,286
Number of PSUs = 449                   Population size = 13,281.871
Design df = 434


0: east = 0
1: east = 1


----------------------------------------------------------------


                           Linearized


Over          |    Mean      Std. Err.    [95% Conf. Interval]


----------------------------------------------------------------



aoj11_recode |
0            |  42.97838  .8478024     41.31207     44.64469
1            |  31.46284  2.57176      26.40819     36.5175


----------------------------------------------------------------
```

While the difference may not be substantively large in this particular example, it does illustrate that using the incorrect weighting variable can lead to different estimates of the quantity of interest. For survey researchers who want to minimize analytical errors, it is important to be mindful of the role and purpose of weighting variables for each analysis.

Notes

1. West, Sakshaug, and Aurelien (2016).

2. Weisberg (2009).

3. For more information on selection of respondents, see https://www.vanderbilt.edu/lapop/ab2018/AmericasBarometer_2018-19_Technical_Report_W_102919.pdf.

4. See Silver et al. (2019) for discussion of this challenge in face-to-face surveys.

5. The appendix table of this note includes a summary of the weights used in each AmericasBarometer study. For further information about each study design, data collection, and information about how to correctly specify weights and declare sampling information for analysis in Stata, see the technical information document that accompanies each study.

6. For more information on how weight1500 is calculated, see https://www.vanderbilt.edu/lapop/docs/AmericasBarometer_weighting_scheme_all_years_of_AB_v2.pdf.

7. An alternative approach would be to weight each country's respondents proportional to their population size. Thus, continuing with the example, responses from Uruguay would be greatly down-weighted compared to Brazilian responses. LAPOP considers countries to be relevant units of analysis and therefore the standard for regional analyses is to weight each country equally.

8. For an example of how using "wt" rather than "weight1500" can influence one's results, see the supplementary appendix material.

9. By complex design, I am referring to features such as stratification and cluster sampling

that distinguish AmericasBarometer samples from simple random samples, which are not feasible in face-to-face surveys. This is addressed using the `svy` command in Stata; see subsequent footnote.

10. These analyses were conducted using Stata 14.2. A design-based analysis of survey data in Stata means using the `svy` prefix with estimation commands (see `help svy_estimation` within Stata for more information).

11. While the stratification used in AmericasBarometer sample selection reduces the sampling variance of estimates, the cluster sampling increases it. The overall effect on variances depends on the particular sample design as well as whether the sample includes weights, which can increase sampling variance.

12. See https://www.stata.com/manuals13/svysvyset.pdf for documentation on `svyset`.

References

Silver, Laura, Martha McRoy, Kat Devlin, and Patrick Moynian. 2019. "Who's Home and Who Isn't? The Challenges of Conducting Face-to-Face Interviews in Jordan." Pew Research Center: Decoded. https://medium.com/pew-research-center-decoded/whos-home-and-who-isn-t-the-challenges-of-conducting-face-to-face-interviews-in-jordan-95a2a7081ded.

Weisberg, Herbert F. 2009. *The Total Survey Error Approach: A Guide to the New Science of Survey Research.* Chicago, IL: University of Chicago Press.

West, Brady T., Joseph W. Sakshaug, and Guy Alain S. Aurelien. 2016. "How Big of a Problem is Analytic Error in Secondary Analyses of Survey Data?" PLOS ONE 11 (6): 1–29.

West, Brady T., Joseph W. Sakshaug, and Yumi Kim. 2017. "Analytic Error as an Important Component of Total Survey Error: Results from a Meta-Analysis." In *Total Survey Error in Practice,* edited by P. Biemer, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, B. West, E. de Leeuw, and N. C. Tucker. Hoboken, NJ: Wiley.

Oscar Castorena (oscar.castorena@vanderbilt.edu) is a Postdoctoral Researcher at LAPOP.

This report was edited by Dr. Elizabeth J. Zechmeister and Dr. Carole J. Wilson. This report was translated by Dr. Juan Camilo Plata, Mariana V. Ramírez Bustamante, and Dr. J. Daniel Montalvo. This report was audited by Mariana V. Ramírez Bustamante. Formatting, production, copy editing, graphics, and report distribution were handled by Rubí Arana and Laura Sellers. Our data and reports are available for free download on the project website. Please follow us on Twitter or Facebook to stay in touch.

As a charter member of the American Association for Public Opinion Research (AAPOR) Transparency Initiative, LAPOP is committed to routine disclosure of our data collection and reporting processes. More information about the AmericasBarometer sample designs can be found at vanderbilt.edu/lapop/core-surveys.

This *Methodological Note* is made possible by the support of the American People through the United States Agency for International Development (USAID) and Vanderbilt University. The contents of this *Methodological Note* report are the sole responsibility of its author and LAPOP and do not necessarily reflect the views of USAID, the United States Government or any other supporting organization. LAPOP's AmericasBarometer surveys are supported predominantly by USAID and Vanderbilt University. The 2018/19 round also had support from the IADB, the UNDP, the Open Society Foundations, and academic partners and researchers across the Americas.

vanderbilt.edu/lapop
@lapop_barometro
@LatinAmericanPublicOpinionProject
lapop@vanderbilt.edu
+1-615-322-4033
230 Appleton Place, PMB 505, Suite 304, Nashville, TN 37203, USA