December 2, 2019

Dear Sam,

I'm working on post-stratification weighting for state weights.[1] What follows is an early writeup that I shared with the CCES Harvard-Tufts team earlier last month.

On the topic of weighting, I (along with Brian and Steve) would like to learn more about how YouGov weights the CCES Common Content. The 2018 CCES Guide writes that, to paraphrase,

> "The completed pre-election interviews are matched to the target frame, using a weighted Euclidean distance metric conditioning on registration status × age × race × gender × education. — (a)
>
> "[Then, the sample] is weighted to adjust for any remaining imbalance that exists among the matched sample. For each team and the common content, the completed cases are weighted to the sampling frame using entropy balancing, ... to match the distributions of the 2017 ACS... The moment conditions included age, gender, education, race, plus their interactions. — (b)
>
> "The resultant weights were then post-stratified by age, gender, education, race, "born again" status, voter registration status, and 2016 Presidential vote choice, as needed. — (c)
>
> "Finally, the weights were post-stratified across states and statewide political races (for governor and senator). — (d)

We have a couple of questions here, including:

1. What is the difference between (b) and (c)? What does it mean for *weights* (instead of respondents) to be post-stratified to state-level outcomes?
2. In the (c) weighting, what is the target value for 2016 Presidential vote choice?
3. How is (d) done exactly, for instance when there are many people who are undecided or did not vote, some states have both Governor and Senator races and others have neither?
4. Are any weights (fractional matches) calculated in (a) and used, for example as starting values, in the second weighting stage?

[1] This work is in collaboration wth Soichiro Yamauchi, who is a political methodology graduate student at Harvard who mainly works on causal inference. I have also benefited with discussions with Steve and the Columbia MRP team (Ben Bales and Lauren Kennedy).

I look forward to discussing — thank you in advance.

Shiro

Survey Weighting for (Moderately Large) Subnational Samples

Shiro Kuriwaki*

December 2019

Abstract

Although researchers routinely used post-stratification weights provided by survey firms, these only weight to the national adult population. When researchers attempt to estimate quantities at the subnational level, they either (a) erroneously apply national weights to subnational subsets of the survey or (b) conduct multi-level regression post-stratification (MRP). Here, using YouGov's weights for the CCES and creating my own state-level weights, I demonstrate how doing the former may skew samples less representative than no weights at all, and I clarify how doing the former latter can replaced by estimating a new set of weights directly targeted towards each geographies. MRP and post-stratification weights have opposite consequences: MRP trades off unbiasedness for lower variance, while weighting trades off efficiency for less bias. While post-stratification weights cannot be used for small groups, they are useful for somewhat larger geographies such as state.

*Ph.D. Candidate, Department of Government and Institute of Quantitative Social Science, Harvard University.
Thanks to Soichiro Yamauchi for helpful discussions.

In this memo I explore how representative CCES samples are at subnational geographies, and whether post-stratification weights ameliorate the problem. Although post-stratification weights are designed to coerce samples to be representative, one set of weights makes a sample representative to only one target, in CCES' case the entire U.S. It is not clear if those same weights *also* make subgroup samples representative of their respective subgroup targets.

It is hard to discuss or propose survey weighting without comparing it to the increasingly popular method of Multi-level Regression Post-stratification (MRP). In many cases, MRP is preferable to weighting because of its flexibility in combining granular post-stratification cells. However, MRP has several limitations relative to a weighting approach. First, often times analysts only have marginal, not joint, distributions as targets, namely partisanship and turnout. Second, one MRP model is specific to one outcome, so projects that involve multiple outcome questions would involve building as many multi-level models as there are outcomes. In contrast, once post-stratification weights are defined, it can be used for any sets of outcomes.

The rest of this memo has several (currently not very well conneected) parts. In Section 1, I compare MRP and weighting. In Section 2, I review how YouGov creates its weights and how I construct state rim weights to mirror that process. In Section 3, I evaluate how YouGov weighting affects the representativeness of different subnational geographies. In Section 4, I do the same evaluation but for rim weights.

1 Post-stratification Weights and MRP

Each method has its pros and cons, summarized by the following table:

	(Unpooled) Weighting	MRP
Applying to different outcomes	Easy, one-stop	Hard, per outcome
Applying to different subgroups	Hard, per geography	Easy, one-stop
Pools respondents from different geographies?	No	Yes
Impose modeling assumptions on...	The selection model	The outcome model
Bias-variance tradeoff	Low bias for more variance	Lower variance for more bias

The first two rows are about practicality. Weights are easy to apply to different outcomes (i.e. questions) because once constructed, researchers can apply it to all questions. In contrast, for MRP requires one model per outcome. For example, there needs to be one multilevel regression model for voting for Presidential vote choice, another for Governor vote choice, and so on. In contrast, weights are tedious to apply for different subgroups. Although it is not well-understood in practice, a single vector of weights cannot simultaneously weight to different targets. For example, if we were to explicit weight the entire CCES to each state-specific distribution, we would need 50 columns of weights, one for each state. In contrast, once an MRP model is constructed for a given outcome, it can be flexibly applied to any geography or subgroup in the geography because it makes predictions for each cell.

The third to fifth in the table addresses the assumptions being made and their consequences. In typical weighting, there is a weight attached to each observation and observations are not pooled (i.e. a respondent in Iowa is not used as a component to estimate outcomes in Wisconsin). This lowers bias because it does not risk carrying unobserved confounding into the estimate, but increases variance. MRP pools respondents, which improves precision (which is a severe problem in small geography) at the cost of making an ignorability-on-observables assumption and risking some unobservable bias due to pooling.

2 Setup and Methodology

The goal of the rest of this memo is to document how weighting affects the distribution of covariates at different sub-geographies. I do not discuss MRP here — this is because the joint distributions of the covariates that are available are matched exactly by construction, and the worry for MRP is about the outcome model, not the weighting model.
In this memo, I consider three types of estimates:

1. Unweighted sample proportions
2. Weighted proportions with YouGov's national post-stratification weights
3. Weighted proportions with custom state-specific weights.

2.1 YouGov's national weights

As a baseline I use the `commonweight` from the 2018 CCES. The guide reports that:

> "The [matched] cases and the frame were combined and the combined cases were balanced on multiple moment conditions using the 2017 ACS. ... First, for the common content, the completed cases were weighted to the sampling frame using entropy balancing. ... The CCES sample was weighted to match the distributions of the 2017 ACS ...
>
> "The moment conditions included age, gender, education, race, plus their interactions. The resultant weights were then post-stratified by age, gender, education, race, "born again" status, voter registration status, and 2016 Presidential vote choice, as needed. Additionally, for the common content, the weights were post-stratified across states and statewide political races (for governor and senator). Weights larger than 15 in the common content were trimmed and the final weights normalized to equal sample size.

Although we do not have access to YouGov's full code, we can partly reproduce this procedure. The ACS provides their own estimates of marginal and some distribution of demographics at the national, state, and congressional district level. We uses those as a source of our target distribution.

2.2 Creating State-specific weights

I created simple rim weights by going state sample by state sample, and assigning a set of weights that targeted marginal distribution of gender, age, and education in that state (as reported by the ACS).

I used the 2017 ACS ($N = 3,210,525$). This includes $267,971$ non-citizens, but we keep them in for now because they are part of the ACS calibrated counts. The covariates are those mentioned at the beginning of this memo: gender, age, race, and education. This writeup uses the state-level counts that ACS provides, although I've also tried using the individual ACS.

We separate states into three tiers based on a data-availability basis:

1. Large states, where all six pairwise interaction of the categories has no zero-cells in the CCES. These include the top seven largest states in the CCES. These states are California, Texas, Florida, New York, Ohio, Pennsylvania, and Illinois. We calculate rim weights by the **marginals and the interactions**.

2. Medium states, where at least pairwise interaction has at least one zero cell, but where all the marginals are populated in the CCES. We calculate rim weights here by the **marginals only**.

3. Small states, where even some states have missing cells. This happens only in race, and those states are Alaska, Delaware, and North Dakota. We calculate the rim weights here by **marginals only, ignoring the zero cell altogether**.

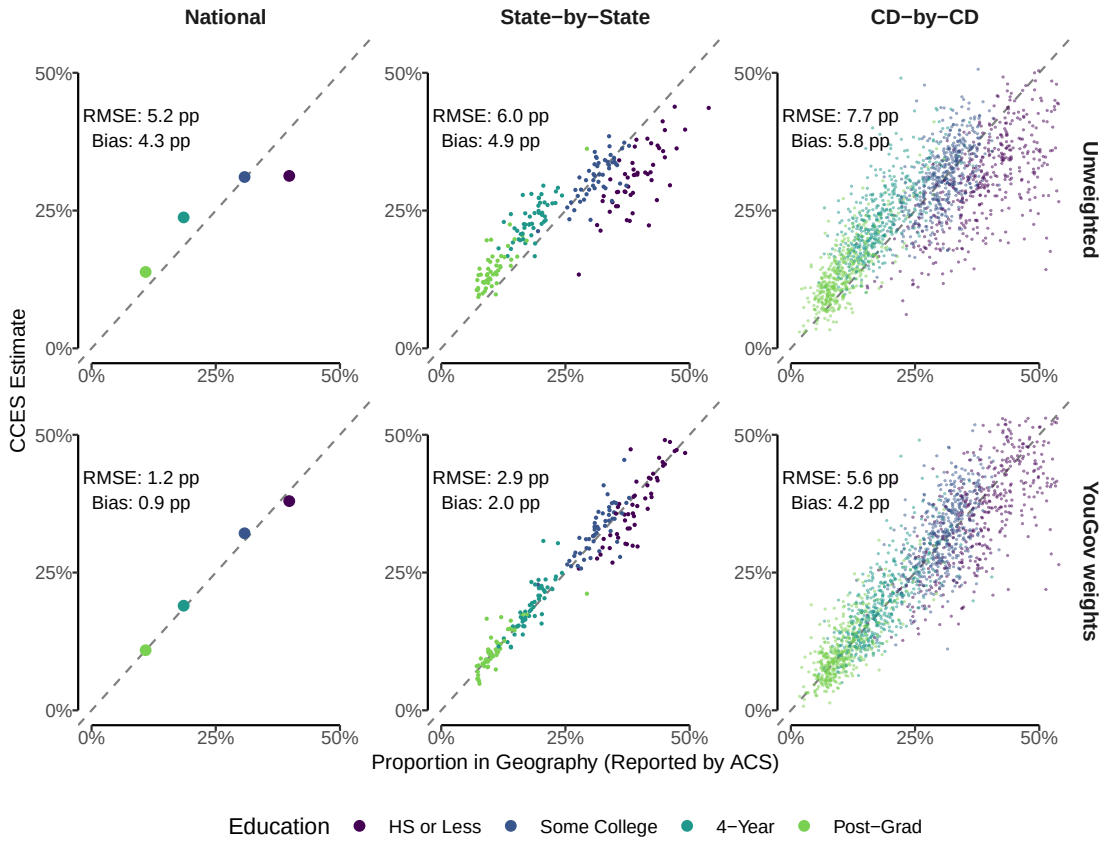
We conducted this using the ACS with sampling weights and the ACS without sampling weights. The resulting rim weights for each state are called `weight_st_wacs` and `weight_st_uacs`, respectively.

3 YouGov Weight Results

We first start by evaluating one metric, education, in Figure 1. We notice several things from the figure:

1. The first set of plots in the first row show that small-samples are on average less representatives than larger ones.
2. The second row, by comparison, shows that YouGov's weights make the estimates more representative. Although the weights primarily target the national distribution, (a) the weighted average for national estimates are not perfect, and (b) state and district estimates are improved as well.
3. Most of the improvement in the second comes from a reduction in bias rather than reduction in variance.
4. There is a smaller reduction in bias in the district level estimates.

Figure 1: Representativeness of samples at different levels of geographies, education

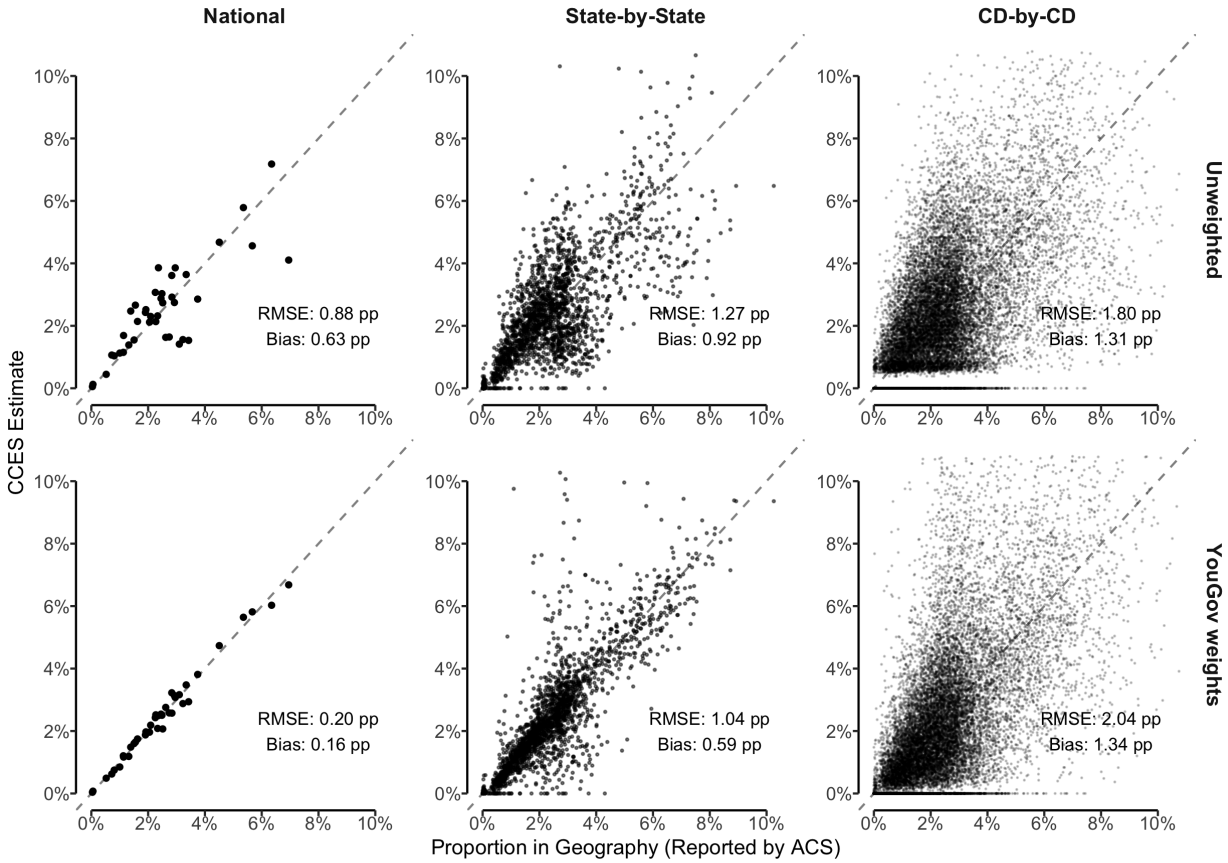


CCES 2018, ACS 1yr 2017. All CCES weighting uses YouGov's national weights, even for state/CD subsets in middle/right panels. CCES/ACS estimate the proportion of an education cell (6 combinations) per geography (1 nation, 50 states, or 435 CDs).

We next take a look not only at education, but the representativeness in terms of education-gender-age joint distributions. YouGov weights on first moments and (presumably two-way) interactions, so their weighted are not guaranteed to hold for three-way joint distributions. Figure 2 shows those quantities of interest, proliferating the number of points to examine. We find:

1. National estimates improve about as much as in Figure 1 in ratio terms.
2. State estimates also improve somewhat, but not by much (8 percent reduction in RMSE, as opposed to 40 percent in the marginal distribution case).
3. District estimates do not improve, and its bias *increases* slightly by 0.08 percentage points. The bias variance decomposition suggests that the variance has increased as well.
4. Some of the outliers in the state estimate suggests that the YouGov national weights up-weight some state-demographic cells in a way that makes them less representative of the state.

Figure 2: Representativeness of samples at different levels of geographies, education × age × gender fraction



Source: CCES 2018, ACS 1yr 2017. All CCES weighting uses YouGov's national weights, even for state/CD subsets in middle/right panels. CCES/ACS estimate the proportion of a {gender x age bin x education} cell (60 combinations) per geography (1 nation, 50 states, or 435 CDs)

4 State Weights Results

Do weights that are specifically targeted to match on the state-specific moments improve representativeness? How does it improve representativeness at levels larger lower than its target? Figure 3 shows figures analogous to the prior figures.

1. As expected, the weights coerce the margins to match the population targets.
2. Although each state subsample's weights is computed separately, its concatenation makes national estimates representatives too.
3. At the smaller district level, the marginal estimates have also improved, and slightly outperform the YouGov weights.
4. However, rim weights do not necessarily improve representativeness of joint distributions. The bottom panel shows that those estimates are about as representative as those with YouGov weights, perhaps by reducing the variance.

Figure 3: Representativeness with custom state-by-state rim weights