Weighting for Survey Experiments

Janet Xu

Soc Stats Reading Group

May 4th, 2017

Developing Standards for Post-Stratification Weighting in Population-Based Survey Experiments

Annie Franco, Neil Malhotra, Gabor Simonovits, and L.J. Zigerell

- Experiments often rely on samples that may be less representative than those recruited by more traditional techniques because the sampling frames contain high coverage error.

Developing Standards for Post-Stratification Weighting in Population-Based Survey Experiments

Annie Franco, Neil Malhotra, Gabor Simonovits, and L.J. Zigerell

- Experiments often rely on samples that may be less representative than those recruited by more traditional techniques because the sampling frames contain high coverage error.
- Non-probability based sampling doesn't interfere with a researcher's ability to estimate **SATE, the sample average treatment effect**

Developing Standards for Post-Stratification Weighting in Population-Based Survey Experiments

Annie Franco, Neil Malhotra, Gabor Simonovits, and L.J. Zigerell

- Experiments often rely on samples that may be less representative than those recruited by more traditional techniques because the sampling frames contain high coverage error.
- Non-probability based sampling doesn't interfere with a researcher's ability to estimate **SATE, the sample average treatment effect**
- Things get harder if the researcher tries to provide an unbiased estimate of the **PATE, the average treatment effect for the corresponding population of interest**

SATE vs. PATE

One of two assumptions need to infer PATE from a given sample:

1. No treatment effect heterogeneity
2. Random sampling of the population

SATE vs. PATE

No Treatment Effect Heterogeneity

Authors seem convinced that this is rare in the social sciences

Random Sampling

- Not true for most experimental samples

SATE vs. PATE

No Treatment Effect Heterogeneity

Authors seem convinced that this is rare in the social sciences

Random Sampling

- Not true for most experimental samples
- Even probability samples cannot be considered truly random given declining of response rates

Using weights to address systematic non-response or self-selection into samples

Core Idea

Use information about the differences between the sample and the population of interest in order to estimate population quantities via adjustment of sample quantities

Even if the sampling probability differs across subgroups, sampling can be assumed to be random within subgroups based on observable covariates (mostly demographics). If this missing-at-random assumption holds, an estimator which weights strata-specific treatment effects by strata-specific inverse response probabilities is unbiased.

But weighted estimates are no longer unbiased estimates of PATE when there exists any unobserved, individual-level factors that is correlated with both treatment effect and sampling probability conditional on observables.

Example

- Effect of treatment is stronger for those with higher interest in politics

But weighted estimates are no longer unbiased estimates of PATE when there exists any unobserved, individual-level factors that is correlated with both treatment effect and sampling probability conditional on observables.

Example

- Effect of treatment is stronger for those with higher interest in politics
- People with higher interest in politics are more likely to self-select into a study or respond to a survey

But weighted estimates are no longer unbiased estimates of PATE when there exists any unobserved, individual-level factors that is correlated with both treatment effect and sampling probability conditional on observables.

Example

- Effect of treatment is stronger for those with higher interest in politics
- People with higher interest in politics are more likely to self-select into a study or respond to a survey
- To recover PATE, we'd need to weigh on political interest

But weighted estimates are no longer unbiased estimates of PATE when there exists any unobserved, individual-level factors that is correlated with both treatment effect and sampling probability conditional on observables.

Example

- Effect of treatment is stronger for those with higher interest in politics
- People with higher interest in politics are more likely to self-select into a study or respond to a survey
- To recover PATE, we'd need to weigh on political interest
- But political interest is not observable in the population of non-respondents!

Practical Problems with Weighting

Covariate Imbalance

Weighting procedures applied to the entire sample (as opposed to within treatment groups) can lead to covariate imbalance across experimental conditions

Loss of Precision

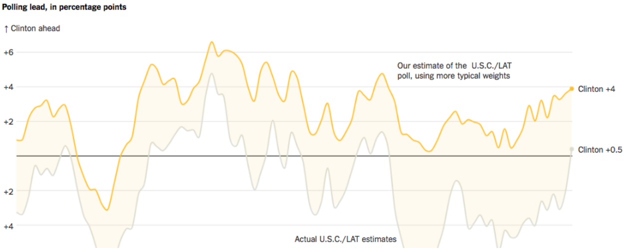
When weights are finer-grained, they make the assumption of equal selection probability within cells more plausible. But that also leads to increased variability in survey weights, which in turn leads to a loss of precision. Weighting also complicates estimation of sampling variance of estimated treatment effects.

THE 2016 RACE

How One 19-Year-Old Illinois Man Is Distorting National Polling Averages

Nate Cohn @Nate_Cohn OCT. 12, 2016

505



Our Trump-supporting friend in Illinois is a surprisingly big part of the reason. In some polls, he's weighted as much as 30 times more than the average respondent, and as much as 300 times more than the least-weighted respondent.

Alone, he has been enough to put Mr. Trump in double digits of support among black voters. He can improve Mr. Trump's margin by 1 point in the survey, even though he is one of around 3,000 panelists.

"The use of weighting methods in published word employing survey experiments is haphazard. Some articles report and discuss only weighted results, while others present only unweighted results. More importantly, most published articles fail to justify this methodological choice (e.g. simply stating in a short footnote that weights were applied). Because reviewers and editors do not seem to require authors to justify the choice of weighting methodology, researchers may cherry-pick estimates based on substantive or statistical significance. **Thus, in current practice weighting is a researcher degree of freedom akin to the selective reporting of outcome variables, experimental conditions, and model specifications.**"

Is this a first-order concern in sociology?

Six Recommendations by Franco et al.

1. Researchers should explicitly state if the quantity of interest is PATE and whether they are reporting unweighted or weighted analyses
2. Researchers should always report SATE
3. If researchers interpret an unweighted finding as a PATE, they should provide evidence that either treatment effect is constant across subgroups or that the sample is a random sample of the population of interest.
4. If researchers interpret a weighted finding as a PATE, then they should be transparent about how the weights were constructed and applied.
5. Researchers using convenience samples should consider constructing weights based on some of the available demographic data for which there is sufficient variance.
6. Report full list of demographic characteristics and benchmark values used to construct the weights.

Worth Weighting? How to Think About and Use Sample Weights in Survey Experiments

SATE vs PATE again

$$\tau_{\mathcal{S}} = \frac{1}{n}\sum_{i\in\mathcal{S}}\Delta_i = \frac{1}{n}\sum_{i\in\mathcal{S}} y_i(1) - \frac{1}{n}\sum_{i\in\mathcal{S}} y_i(0).$$

$$\tau = \frac{1}{N}\sum_{i=1}^{N}\Delta_i = \frac{1}{N}\sum_{i=1}^{N} y_i(1) - \frac{1}{N}\sum_{i=1}^{N} y_i(0).$$

Sampling Probability for unit i

$$\pi_i \equiv \mathbf{P}\{S_i = 1\} = \mathbb{E}[S_i]\,,$$

Horvitz-Thompson estimator

$$\hat{y}_{HT} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\pi_i}S_i y_i.$$

Let $\bar{\pi} = \frac{1}{N}\sum \pi_i$ be the average sampling probability

$n = \sum S_i$ be the sample size

$$\mathbb{E}[n] = N\bar{\pi}.$$

let $w_i = \bar{\pi}/\pi_i$ be the *sampling weight*.

Total weight of sample

$$Z \equiv \sum_{i=1}^{N} \frac{\bar{\pi}}{\pi_i} S_i = \sum_{i=1}^{N} w_i S_i.$$

Rewrite Horvitz-Thompson as

$$\hat{y}_{HT} = \frac{1}{\mathbb{E}[n]} \sum_{i=1}^{N} \frac{\bar{\pi}}{\pi_i} S_i y_i = \frac{1}{\mathbb{E}[Z]} \sum_{i=1}^{N} S_i w_i y_i.$$

Hàjek estimator

$$\hat{y}_H = \frac{1}{Z} \sum_{i=1}^{N} w_i S_i y_i$$

Hàjek estimator is not unbiased, but the bias will tend to be negligible by the following lemma

$$\mathbb{E}[\hat{y}_H] - \mu \doteq -\frac{1}{\mathbb{E}[n]}\left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)\frac{\bar{\pi}}{\pi_i}\right) = -\frac{1}{\mathbb{E}[n]}Cov[y_i, w_i]\,.$$

Estimating PATE

First step: estimate sample dependent ν_S

$$\Delta_i = y_i(1) - y_i(0) \text{ for } i \in \mathcal{S}.$$

$$\nu_S = \frac{1}{Z}\sum S_i \frac{\bar{\pi}}{\pi_i}\Delta_i = \frac{1}{Z}\sum S_i \frac{\bar{\pi}}{\pi_i} y_i(1) - \frac{1}{Z}\sum S_i \frac{\bar{\pi}}{\pi_i} y_i(0).$$

Estimating PATE

Second step: estimate population τ

$$\hat{\tau}_{hh} = \frac{1}{Z_1}\sum_{i=1}^{N} S_i T_i \frac{\bar{\pi}}{\pi_i} y_i(1) - \frac{1}{Z_0}\sum_{i=1}^{N} S_i (1 - T_i) \frac{\bar{\pi}}{\pi_i} y_i(0)$$

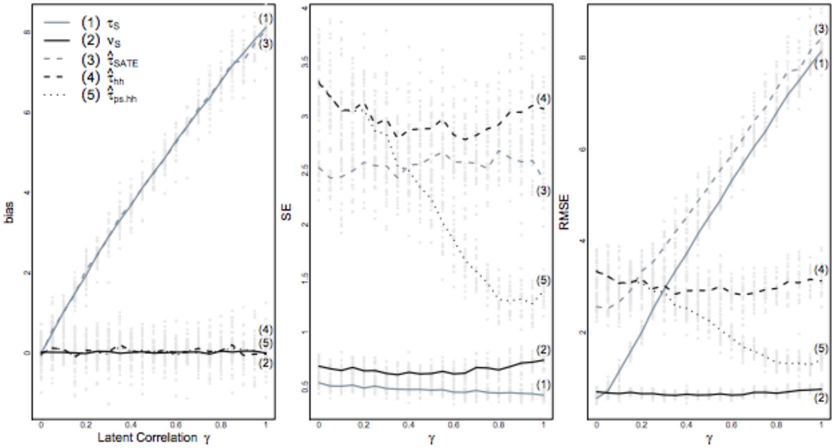
with

$$Z_1 = \sum_{i=1}^{N} S_i T_i \frac{\bar{\pi}}{\pi_i} \text{ and } Z_0 = \sum_{i=1}^{N} S_i (1 - T_i) \frac{\bar{\pi}}{\pi_i}.$$

Simulations

	Estimator	Mean	Bias	SE	RMSE	boot SE	Coverage
A							
1	τ_S	40.36	7.77	1.35	7.89		
2	ν_S	32.58	0.00	1.84	1.84		
3	$\hat{\tau}_{SATE}$	40.37	7.78	3.14	8.39	3.12	30%
4	$\hat{\tau}_{hh}$	32.62	0.03	3.91	3.91	3.87	95%
5	$\hat{\tau}_{ps}$	32.60	0.01	2.67	2.67	2.69	95%
B							
1	τ_S	30.00	0.00	0.00	0.00		
2	ν_S	30.00	0.00	0.00	0.00		
3	$\hat{\tau}_{SATE}$	30.01	0.01	2.58	2.58	2.58	95%
4	$\hat{\tau}_{hh}$	30.03	0.03	3.35	3.35	3.31	95%
5	$\hat{\tau}_{ps}$	30.02	0.02	3.32	3.32	3.29	95%

Simulations



Real Data Application