Survey Research and Design

Survey Weighting

William Marble

October 3, 2023

Last Time

Models of nonresponse:

- Separate cause model (ignorable nonresponse)
- Common cause model (conditionally ignorable nonresponse)
- Survey variable cause model (nonignorable nonresponse)

Last Time

Models of nonresponse:

- Separate cause model (ignorable nonresponse)
- Common cause model (conditionally ignorable nonresponse)
- Survey variable cause model (nonignorable nonresponse)

Weighting intuition:

- Upweight respondents who are underrepresented in the sample relative to the population
- Force the distribution of weighting variables in the sample to match the known population targets

Choosing Weighting Variables

What Variables to Weight On?

- How to determine "underrepresented"? Try to come up with weighting variables that satisfy conditional ignorability
- Must affect both response probability and responses to question of interest
- Must have population targets for weighting variables
- Easy to test over/underrepresentation and whether weighting variables are correlated with response
- Key assumption (**untestable!**): we're adjusting for **all** variables that affect both nonresponse + attitudes

Before Weighting

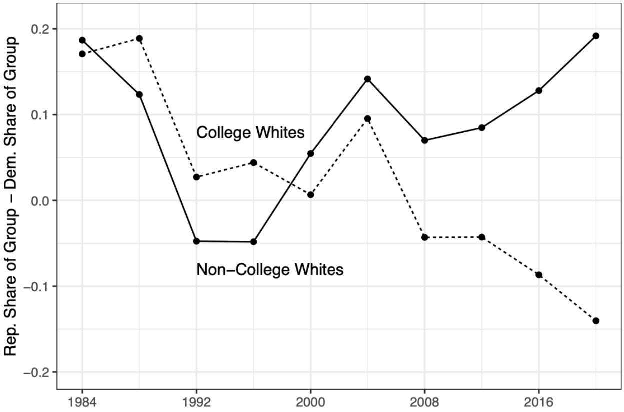
- Do everything possible to ensure match between sample and population on key variables
- We don't want to rely too much on weights: large weights $\rightsquigarrow$ higher uncertainty/standard error
- Potentially oversample subgroups we think will be underrepresented
- Quota sampling (once frowned upon, now common) tries to minimize need for weighting
- Might oversample rural areas or Republicans/Independents (who are less likely to take surveys)

What Variables to Use in Weighting?

- Typically use sociodemographic variables: age, race, sex
- Nowadays also weight on education
- Income?
- Party ID? Past vote choice?
- Region/state?

Growing Educational Polarization

Figure 1: Net Republican Votes in Presidential Races Among Whites, By Education

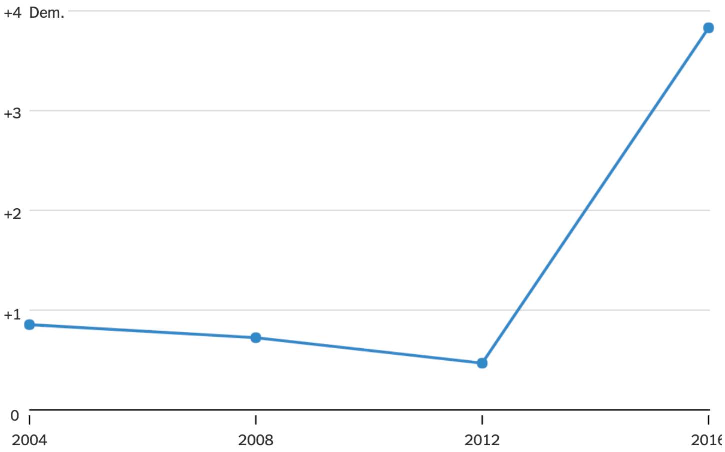


Education Weighting

Education Weighting Vastly More Important in ’16

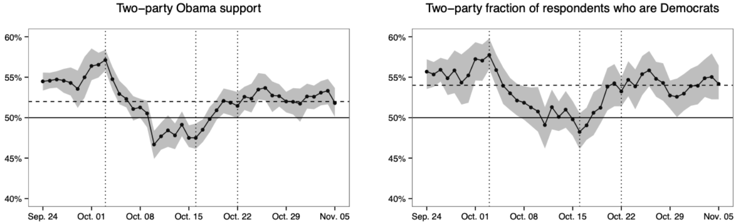
A poll that didn’t weight by education might have been imperceptibly more Democratic-leaning in past elections, but was notably biased in 2016.

The effect of neglecting to weight by education in a typical national survey (pct. margin)

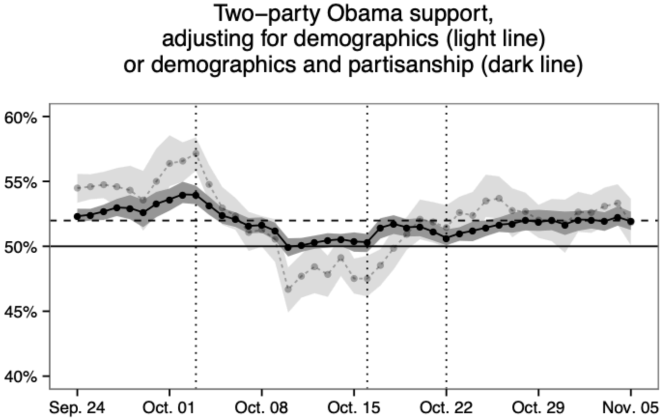


Weighting on Partisanship

- Partisanship isn't quite the same as age/race/sex/education
- No Census data on partisanship, but party registration data from voter files
- Evidence of over-time differential nonresponse (graph from 2012):



Weighting on Partisanship



The Hows of Survey Weighting

Post-stratification

Recall: Stratified random sample is a random sample within strata defined by the researcher.

Post-stratification

Recall: Stratified random sample is a random sample within strata defined by the researcher.

Post-stratification treats our data *as if* they were generated by a SRS.

From Last Time...

Post-Stratifying on Gender

Gender	Sample Partisanship Democrat	Republican	Pop. Proportion
Men	40%	60%	50%
Women	60%	40%	50%

Basic idea: reweight sample according to population proportion

More Variables for Post-Stratification

Post-Stratifying on Gender and Age

Gender	Age	Democrat	Republican	Pop. Proportion
Men	18-29	60%	40%	15%
Women	18-29	70%	30%	15%
Men	30-39	45%	55%	12%
Women	30-39	56%	44%	13%
⋮	⋮	⋮	⋮	⋮

Even More Variables...

State $\times$ Age $\times$ Race $\times$ Gender $\times$ Education

51 states (+ DC) $\times$ 6 age groups $\times$ 5 racial groups $\times$ 2 gender groups $\times$ 3 education levels = 9,180 poststratification cells

Most surveys don't have that many respondents $\rightsquigarrow$ poststratification infeasible

Matching Population Margins

Instead of matching the **joint distribution** of the weighting variables, we can match the **marginal distribution**

Matching Population Margins

Instead of matching the **joint distribution** of the weighting variables, we can match the **marginal distribution**

Joint distribution of age $\times$ gender looks like:

	Men	Women
18-40	x_1%	x_2%
41-64	x_3%	x_4%
65+	x_5%	x_6%

with $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 100$.

Matching Population Margins

Instead of matching the **joint distribution** of the weighting variables, we can match the **marginal distribution**

Joint distribution of age $\times$ gender looks like:

	Men	Women
18-40	$x_1\%$	$x_2\%$
41-64	$x_3\%$	$x_4\%$
65+	$x_5\%$	$x_6\%$

with $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 100$.

Marginal distributions of age and gender (*separately*) looks like this:

Category	Percent of Population
Men	$x_1\%$
Women	$x_2\%$
18-40	$y_1\%$
41-64	$y_2\%$
65+	$y_3\%$

with $x_1 + x_2 = 100$ and $y_1 + y_2 + y_3 = 100$

Raking

- Most commonly used weighting method is called “raking” or “iterative proportional fitting”
- Ensures matches on specified margins (but not full joint distribution)
- Tends to generate weights that are similar to weights you get from other methods

Raking Algorithm

Suppose we are weighting on K variables. Start with weights of 1 for each unit in sample. (Or start with design-based weights.)

Raking Algorithm

Suppose we are weighting on K variables. Start with weights of 1 for each unit in sample. (Or start with design-based weights.)

Then repeat the following steps over and over until the results don't change...

1. Update weights to match margins on Variable 1
2. Update weights to match margins on Variable 2

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

K. Update weights to match margins on Variable K

Raking Algorithm

Suppose we are weighting on K variables. Start with weights of 1 for each unit in sample. (Or start with design-based weights.)

Then repeat the following steps over and over until the results don't change...

1. Update weights to match margins on Variable 1
2. Update weights to match margins on Variable 2

$\vdots$ $\qquad\qquad$ $\vdots$

K. Update weights to match margins on Variable K

For each step, we compute the weight updates as $\frac{\text{Population Proportion}}{\text{Weighted Sample Proportion}}$, then multiply the old weights by the updates to obtain new weights.

Implementing raking algorithm $\rightsquigarrow$ `R` code

Trimming Weights

- More weighting variables $\rightsquigarrow$ high-variance weights and more large weight
- After weighting, may want to "trim" very small and very large weights

Trimming Weights

- More weighting variables $\rightsquigarrow$ high-variance weights and more large weight
- After weighting, may want to "trim" very small and very large weights
- Large weights $\rightsquigarrow$ individual people are very influential:

TheUpshot

THE 2016 RACE

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

Trimming Weights

- More weighting variables $\rightsquigarrow$ high-variance weights and more large weight
- After weighting, may want to "trim" very small and very large weights
- Large weights $\rightsquigarrow$ individual people are very influential:

  TheUpshot

  THE 2016 RACE

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

- Often trim to be between $[0.2, 4]$, but also see stuff like $[0.1, 7]$. No single right answer.

Accounting for Weights in Standard Error Calculation

Weighting reduces the "effective sample size"

Intuition: people with large weights have a lot of influence $\rightsquigarrow$ higher variance if that person was/wasn't in the sample

Account for this in standard error/margin of error using "design effect" (*deff*):

$$SE_{wtd} = SE_{unwtd} \times \sqrt{deff}$$

$$deff = 1 + \left(\frac{\text{sd(weights)}}{\text{mean(weights)}}\right)^2$$

Effects of Weighting in October 2022 PORES/SurveyMonkey Poll