GV205
Measuring Public Opinion
Imperfect Samples

05 March, 2018

Joe Greenwood
YouGov
@NiceOneCombo

Overview

- Approaches to Inference
  - Design-Based
  - Model-Based
- A Key Assumption: Ignorability
- Weighting:
  - A Simple Example
  - Another Example
  - In Practice
  - Adjustments
- Summary

Two Approaches to Inference

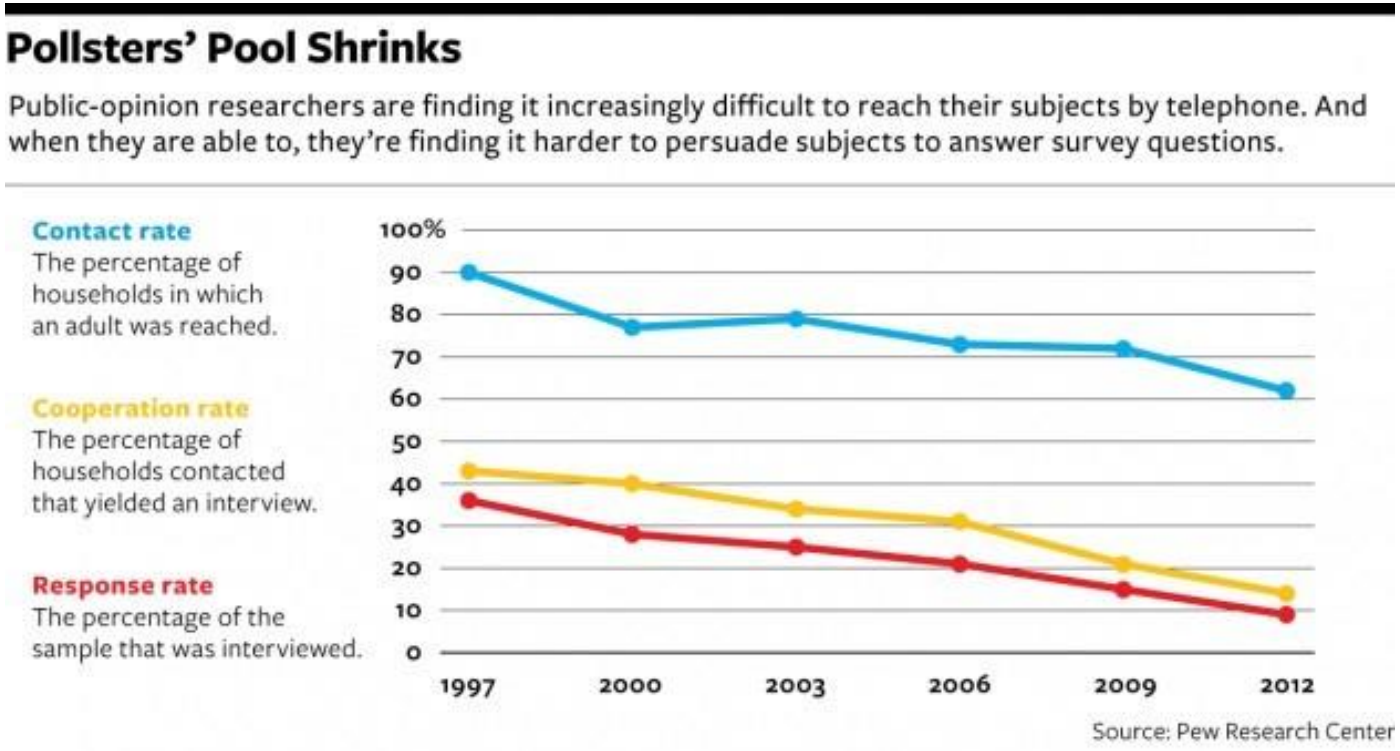
- Design-based inference (random sampling)
  - The inferences that we make on the basis of the data rest on the internal design of the study, such as random selection of cases or (as will be considered next week) random allocation to experimental conditions

- Model-based inference (weighting)
  - The inferences that we make on the basis of the data rest on external information not associated with the design of the study, such as the representativeness of the data based on key population measures

Design-Based Inference: A Key Assumption

- Ignorability for the design-based (random sampling) approach:
  - Response and non-response are unrelated to the variables of interest
  - For example, we assume that those who respond are no more or less likely to support a given policy (if they are then there's a problem)

The Problem



http://blog.qsample.com/the-looming-online-survey-crisis-and-one-way-to-avoid-it/

The Problem

- In practice there is no perfect random survey sample
- Response rates (to ‘random’ sample surveys) are low and declining:
  - ‘At response rates of 10% or 20%, statistical theory no longer provides a basis for generalizing from sample to population.’

Pasek, p. 18

A Solution

- Since the design-based approach (i.e. random sampling) to survey inference is becoming more difficult…
- …we need to consider the model based approach (i.e. weighting).
- But, weighting is not the golden bullet; it can only do so much.

Model-Based Approach: A Key Assumption

- Ignorability for the model-based (weighting) approach:
  - Response and non-response are unrelated to the variables of interest, conditional on key covariates
  - For example, we assume that women who respond are no more or less likely to support a policy than are women who do not respond, even if women are more or less likely to respond overall, and we can weight the sample to ensure that it has a representative proportion of women (based on what we know about the population)

Weighting: A Simple Example

- Suppose that we know that the population has approximately 51% women and 49% men
- Now suppose that we obtain a sample with 56% women and 44% men: what weights do we need to apply to our sample?
- Why is this important?

Weighting: A Simple Example

- For women in the sample: (1/56) x 51 = 0.91
- For men in the sample: (1/44) x 49 = 1.11
- So, if we had 560 women in the sample then they come to represent 560 x 0.91 = 510 (with rounding) women
- And, if we had 440 men in the sample then they come to represent 440 x 1.11 = 490 (with rounding) men
- Thus, women are downweighted and men are upweighted so that the sample reflects what we know about the population
- Again, why is this important?

Weighting: Another Example

- We choose weights such that the group shares are identical to the shares in the target population
- This also works for variables with more than two groups, such as categorised age (e.g. 18-34, 35-49, 50-64, 65+)
- We can also use weights for more than one variable that segments the population at a time…

Weighting: Another Example

- In the population:

Gender	Age				Total
	18-24	25-49	50-65	65+	
Female	5.7%	21.5%	12.0%	11.6%	50.7%
Male	5.5%	20.9%	11.6%	11.3%	49.3%
Total	11.2%	42.4%	23.6%	22.9%	100%

Weighting: Another Example

- In our hypothetical sample:

Gender	Age				Total
	18-24	25-49	50-65	65+	
Female	4.5%	22.1%	14.7%	14.4%	55.8%
Male	3.6%	17.6%	11.7%	11.4%	44.2%
Total	8.1%	39.7%	26.4%	25.8%	100%

Weighting: Another Example

- To calculate our weights:

Gender	Age			
	18-24	25-49	50-65	65+
Female	(1/4.5) x 5.7 = 1.27	(1/22.1) x 21.5 = 0.97	(1/14.7) x 12.0 = 0.82	(1/14.4) x 11.6 = 0.81
Male	(1/3.6) x 5.5 = 1.53	(1/17.6) x 20.9 = 1.19	(1/11.7) x 11.6 = 0.99	(1/11.4) x 11.3 = 0.99

Weighting: Another Example

- In numbers:

Gender	Age				Total
	18-24	25-49	50-65	65+	
Female	45 x 1.27 = 57	221 x 0.97 = 215	147 x 0.82 = 120	144 x 0.81 = 116	508
Male	36 x 1.53 = 55	176 x 1.19 = 209	120 x 0.99 = 116	116 x 0.99 = 113	493
Total	112	424	236	229	1,001

Weighting: In Practice

- At YouGov, we quota and weight by:
  - Age
  - Gender
  - Education
  - Social grade
  - Region
  - 2017 general election vote
  - 2016 EU referendum vote
  - Political attention
- Some of the above are ‘interlocking’.

Weighting: In Practice

Sex	Population (ONS)	Sample (BES)	Difference	Weighted (BES)	Difference
Male	49.3	45.6	-3.7	49.0	**-0.3**
Female	50.7	54.4	3.7	51.0	**0.3**

Weighting: In Practice

Age Group	Population (ONS)	Sample (BES)	Difference	Weighted (BES)	Difference
18-24	11.2	7.2	-4.0	11.0	**-0.2**
25-49	42.2	36.3	-5.9	38.8	**-3.4**
50-65	23.6	28.2	4.6	27.4	**3.8**
65+	22.9	28.3	5.4	22.9	**0.0**

Weighting: In Practice

Education Level	Population (ONS)	Sample (BES)	Difference	Weighted (BES)	Difference
High	31.0	41.4	10.4	35.3	**4.3**
Medium	39.4	31.3	-8.1	38.0	**-1.4**
Low	29.6	27.3	-2.3	26.6	**-3.0**

Weighting: In Practice

Region	Population (ONS)	Sample (BES)	Difference	Weighted (BES)	Difference
South	32.4	30.8	-1.6	32.9	**0.5**
London	13.4	9.7	-3.7	11.6	**-1.8**
Midlands	16.5	17.5	1.0	16.6	**0.1**
North	24.1	27.5	3.4	24.7	**0.6**
Scotland	8.7	8.7	0.0	8.9	**0.2**
Wales	4.9	5.9	1.0	5.2	**0.3**

Weighting: In Practice

Party	Population (BBC)	Sample (BES)	Difference	Weighted (BES)	Difference
Conservative	29.8	32.0	2.2	29.7	-0.1
Labour	28.2	34.6	6.5	32.7	4.5
Liberal Democrat	5.2	5.6	0.4	5.5	0.3
Green	1.1	1.5	0.4	1.0	-0.1
UKIP	1.3	1.7	0.5	1.9	0.6
SNP	2.1	2.4	0.3	2.1	0.0
Plaid Cymru	0.4	1.1	0.8	0.8	0.4
Other	0.6	0.2	-0.5	0.1	**-0.5**
Did not vote	31.3	20.9	-10.4	26.2	-5.1

Weighting: In Practice

Campaign	Population (BBC)	Sample (BES)	Difference	Weighted (BES)	Difference
Remain	34.7	41.1	6.4	40.8	**6.1**
Leave	37.5	40.7	3.2	39.4	**1.9**
Did not vote	27.8	18.2	-9.6	19.9	**-7.9**

Weighting: Adjustments

- When we create weights by population targets after we have gathered our data it is known as post-stratification

- Part of the post-stratification process is raking (or Rim Weighting):
  - Raking is an iterative process in which weights are calculated to bring the sample into line with one set of population targets, then adjusted to meet additional targets, and the cycle is repeated until the weights bring the sample as close as possible to as many of the population targets as possible

Weighting: Adjustments

- If we have large weights then we are assuming that those respondents represent many more people in the population
- That assumption is uncertain though; what if the larger group in the population has more variation than the smaller group in the sample?
- This requires us to calculate an Effective Sample Size, which reduces the sample size and thus increases the uncertainty around any subsequent estimates that we make:
  - $Effective\ Sample\ size = \frac{(\sum W_i)^2}{\sum {W_i}^2}$

Weighting: Adjustments

- Using the previous, simple example of a sample with the wrong proportions of men and women:
  - We had 560 women with a weight of 0.91 each, and 440 men with a weight of 1.11 each
  - $Effective\ Sample\ Size = \frac{((560 \times 0.91)+(440 \times 1.11))^2}{\left((0.91^2)\times 560\right)+\left((1.11^2)\times 440\right)}$
  - $Effective\ Sample\ Size = \frac{996{,}004}{1005.86} = 990$

Weighting: Summary

- All surveys nowadays have weights:
  - Even the design-based approach "borrows" from the model-based approach
- Using one variable we can easily create weights, when using more than one we use post-stratification to create weights
- A core assumption in all these approaches is ignorability:
  - Conditional on the stratifying variables (e.g. age, gender, education) there is no difference in the variable of interest between respondents and non-respondents:
    - For example, the 18-24 old men with a university degree that participated are identical to those who did not participate in terms of our variable of interest
- Weighting can only take us so far; it does not correct a bad sample