Sampling Simulation

Shiro Kuriwaki

June 11, 2020

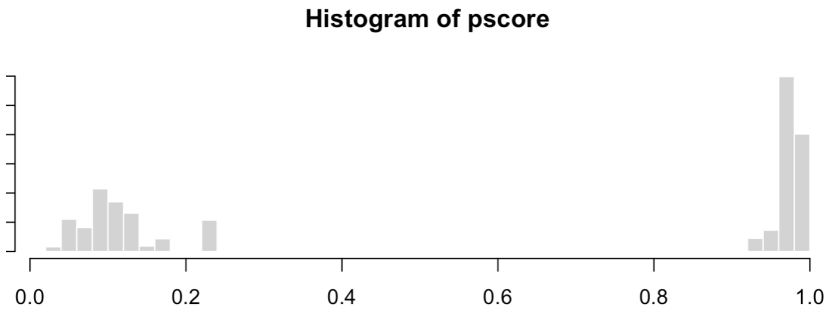
Realistic Simulation, with full control over the sampling scheme

- Population ($N = 600,000$): CCES Data, expanded using post-stratification weights
  - this is technically not a census, but it has a natural covariance structure and makes the simulation realistic
- Sample ($n = 1,000$): Simple Random Sample (SRS), OR a biased sample where the propensity score for population member $i \in \{1, ..., N\}$ is determined by:

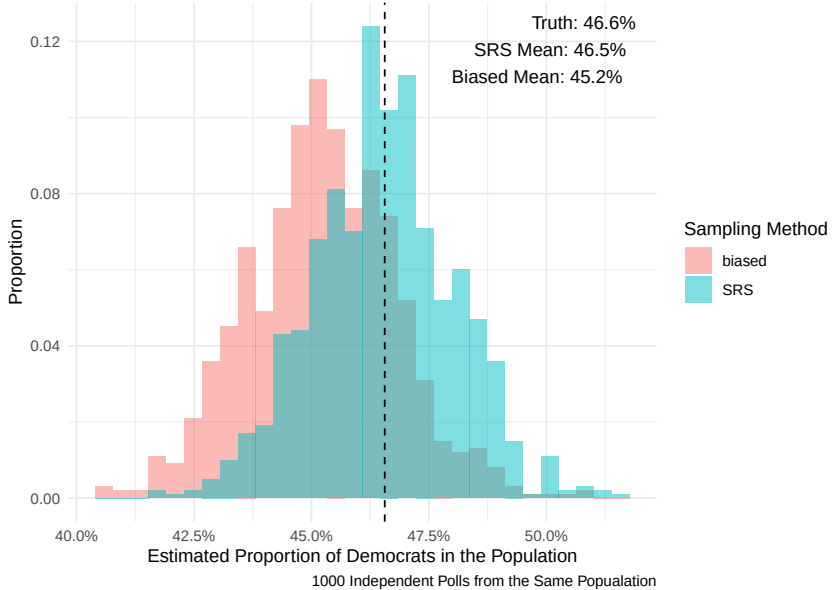
$$= \text{invlogit}\left\{ -4 + \begin{pmatrix} 1.0 \\ 0.8 \\ 0.7 \\ 0.6 \\ 0.5 \end{pmatrix}^\top \begin{pmatrix} \text{White}_i \\ \text{Black}_i \\ \text{Hispanic}_i \\ \text{Asian}_i \\ \text{All Other}_i \end{pmatrix} + \begin{pmatrix} 5.0 \\ 4.0 \\ 1.2 \\ 0.5 \end{pmatrix}^\top \begin{pmatrix} \text{Post-Grad}_i \\ \text{4-Year}_i \\ \text{Some College}_i \\ \text{HS or Less}_i \end{pmatrix} + \begin{pmatrix} 6.0 \\ 1.0 \\ 0.4 \\ 0.3 \end{pmatrix}^\top \begin{pmatrix} \text{Follow News}_i \\ \text{Sometimes}_i \\ \text{Now and Then}_i \\ \text{Hardly}_i \end{pmatrix} \right\}$$

where White_i for example is an indicator variable for whether respondent i is White.

Then I `sample(1:N, size = n, replace = FALSE, prob =` Propensity Score$_i$`)`


Histogram of pscore
0.0
0.2
0.4
0.6
0.8
1.0

Biased Sample Gives Biased Sampling Distribution



1000 Independent Polls from the Same Popualation