

MRP Methodology: Estimating Favorability Ratings for Trump & Clinton

James Wyatt

June 14, 2016

General Overview

Morning Consult asked more than 39,000 registered voters nationally since April 1st, 2016, if they have a favorable or unfavorable opinion of Hillary Clinton and Donald Trump. We use a statistical technique known as multilevel regression and poststratification (MRP) to construct state-level estimates from national polling data.

Data Construction and Specifications

Voters were asked to rate their overall favorability for each candidate via the following question:

Next we will look at a list of names that have been talked about as potential candidates for President. It is a long list, please take the time to go through the list carefully and give an individual answer for each name below. For each person, please indicate if you have a Very Favorable, Somewhat Favorable, Somewhat Unfavorable, or Very Unfavorable opinion of each individual. If you have heard of the person, but do not have an opinion, please mark Heard Of, No Opinion (Donald Trump / Hillary Clinton)

The response options were then combined and recoded to produce our main outcome variable (otherwise known as the dependent variable) coded as follows: for “Total Favorable” and “Total Unfavorable” we combined the proportions of those that selected “Very/Somewhat” for the favorable and unfavorable answer options. For the “Don’t Knows” we merged the proportions of those that selected “heard of, no opinion” or “never heard of”.

Model Fit

After we convert our dependent variable into binary response choices such as 1s (Favorable) and 0s (both Unfavorable and Don’t Know), we fit a three separate multinomial mixed effects logistic regression models (varying intercept models) for each answer option to estimate Trump and Clinton’s favorability rating, which is modeled as a function of each individual’s age, gender, race, education, state, and 2012 presidential election returns in their respective state.

To estimate state-level favorability ratings, we calculate a weighted sum of the individual demographic-geographic type for each state. Namely, we poststratify the predictions from our regression models on age, gender, education and race obtained from the 5-year estimates of the adult citizen population from the 2013 American Community Survey (ACS).

The full model can be defined:

Let \hat{Y} be defined as the predicted probability that any one individual will support Hillary Clinton

$$Pr(y_i = 1) = \hat{Y} = \text{logit}^{-1}(B_0 + \alpha_{k[i]}^{age} + \alpha_{l[i]}^{education} + \alpha_{m[i]}^{gender} + \alpha_{r[i]}^{race} + \alpha_{s[i]}^{state} + Bx_i)$$

Where:

$$\text{logit}^{-1}(\alpha) = \frac{\exp(\alpha)}{\exp(\alpha) + 1}$$

The fixed-effects covariate is:

$$Bx_i = x_{2012_presidential_election_returns}$$

The modeled effects of individual level groups are drawn from a normal distributions with a mean of zero and variance that's unique to each covariate.

$$\begin{aligned}\alpha_k^{age} &\sim N(0, \sigma_{age}^2), \text{ for } k = 1, \dots, 4 \\ \alpha_l^{education} &\sim N(0, \sigma_{education}^2), \text{ for } l = 1, \dots, 4 \\ \alpha_m^{gender} &\sim N(0, \sigma_{gender}^2), \text{ for } m = 1, \dots, 2 \\ \alpha_r^{race} &\sim N(0, \sigma_{race}^2), \text{ for } r = 1, \dots, 3 \\ \alpha_s^{state} &\sim N(0, \sigma_{state}^2), \text{ for } s = 1, \dots, 51\end{aligned}$$

Once we have \hat{Y} for each individual strata based on age, gender, education, race and state, we then compute the poststratified averages of \hat{Y} to develop state estimates of electoral support for each candidate. Specifically, the mixed-effects model develops estimates for each “type” of person and those estimates (favorability for Trump/Hillary) are then weighted in proportion to the frequencies of those types of people in each state.

$$\widehat{MRP}_{state_i} = \frac{\sum_{c \in s} N_c \theta_c}{\sum_{c \in s} N_c}$$