Multilevel Logistic Regression

538 members of the Electoral College

270 electoral votes needed to win

Nominee	Hillary Clinton	Donald Trump
Party	Democratic	Republican
Home state	New York	New York
Running mate	Tim Kaine	Mike Pence

STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

US election

- Multiple polls are conducted nationwide
  - Sample question: will you vote Republican or Democrat?
- Want to estimate the vote by each demographic, in each state
  - 50 states
  - Demographics considered: sex and race
    - E.g., want to know the percent of black men in Nebraska who will vote Republican
  - Lots of demographics, 50 states (+ Washington, DC), and sample size of about 1000
    - Possibly *no* black men in Nebraska were polled at all

Application

- We want to estimate the number of people who will vote Republican
- Do this using *poststratification:* we know how many black men there are in Nebraska, so if we can estimate the percent of black men in Nebraska who will vote Republican, we can add that number to the total
  - Do that with all the demograhpics

Model

$$P(y_i = 1) = logistic\left(\alpha_{j[i]}^{state} + \beta_{female} I_{female,i} + \beta_{black}\, I_{black,i}\right)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_{state}^2)$$

- $y_i = 1$ means the i-th respondent votes Republican
- α_j^{state} are the log-odds of voting republican for a non-black man in state j

Fitting the model

- (In R)
- $\sigma_{state} = 0.4$
  - The log-odds of voting Republican can easily differ by 1.0 in different states
  - ```
    > logit(.4)
    [1] -0.4054651
    > logit(.6)
    [1] 0.4054651
    ```

    Makes sense: some states tend to vote Dem., and some states tend to vote Republican
    ```

A more complicated model

$$P(y_i = 1) = logistic\left(\begin{matrix}\alpha_{j[i]}^{state} + \alpha_{k[i]}^{age} + \beta_{female} I_{female,i} \\ +\beta_{black} I_{black,i} + \beta_{black,female} I_{black,i} I_{female,i}\end{matrix}\right)$$

$$\alpha_j^{state} \sim N(\mu_{\alpha^{state}}, \sigma_{state}^2)$$

$$\alpha_k^{age} \sim N(\mu_{\alpha^{age}}, \sigma_{age}^2)$$

- Added an interaction, and now modelling age (as a categorical variable) as well
- (in R)
- $\sigma_{age} \approx 0.1$, much smaller than σ_{state}
  - Average political leanings vary by state much more than by age