Model Specs—3 Weeks Out

Theodore Dounias

10/2/2018

County Level Models

Model 1 is a fixed-effects, bare-bones model that exclusivelly includes percentage of VBM votes, and dummy variables for year, election type, and county,. Its call would look a bit like:

$$turnout_{c,tp,y} \sim (mail\%)_{c,tp,y} + (C \times B_c) + (T \times B_{tp}) + (Y \times B_y)$$

Where C, T, Y are vectors of dummy variables and c, tp, y are indicators of county, election type, and year. mail% is the percentage of the vote conducted through VBM. This model serves as a baseline for comparing others.

Model 2 A more informed baseline, model 1 plus variables of urban and white population:

$$turnout_{c,tp,y} \sim (mail\%)_{c,tp,y} + (C \times B_c) + (T \times B_{tp}) + (Y \times B_y) + (urban\%)_c + (white\%)_c$$

This would be the "individual" level model from Gelman and Hill. I'm unsure what the "group" level for county would be. Maybe that part of the book would be more helpful for discerning effects on people's individual p-vote?

Maybe more informative is what I did with exercise 12.2. The model tries to predict the concentration of a particular chemical based on treatment of children across time. Therefore the two levels are a visit by one individual child (here an election! so type, vbm_pct, year) and predictors for that individual child that are stable across time, like treatment type, or demographics (here race and urban pop per county).

This means I can fit a model only based on election facts, with a variable for county (models 1,3) or one that takes into account stable characteristics of the county (models 2, 4).

Model 3 A mixed-effects version of model 1, just adds mixed effects for county:

$$turnout_{c,tp,y} \sim (mail\%)_{c,tp,y} + (T \times B_{tp}) + (Y \times B_y) + \alpha_{[c]}$$

$$\alpha_c \sim N(0, \sigma_\alpha^2)$$

Model 4 A mixed-effects version of model 2:

$$turnout_{c,tp,y} \sim (mail\%)_{c,tp,y} + (T \times B_{tp}) + (Y \times B_y) + \alpha_{[c]}$$

$$\alpha_c \sim N(D_c\gamma, \sigma_\alpha^2), \ for \ c = 1, ..., 64$$

Where D is a 2 x 64 matrix of the county level predictors and γ a vector of coefficients for the county-level regression.

Model 5 During one of my discussions with Andrew, we discussed the possibility of making a model that answers the question: "Does VBM affect counties with some particular characteristic *for which I don't have data* more than others?" As such, this model would substitute county-level effects with a set of 3-4 dummy variables created through my intuitive understanding of Colorado politics and counties. For example, maybe

a distinciton between central Colorado urban counties, East Colorado plains counties, and West Colorado mountain counties. The model would look a bit like:

$$turnout_{c,tp,y} \sim (mail\%)_{c,tp,y} + (T \times B_{tp}) + (Y \times B_y) + (my\ county\ type)_i \times (\beta_i)$$

Where i is an indicator between the types of county.

Model 6 As a check on the previous model, run a Principle Components Analysis on full demographic data from the 2010 census, to classify counties in the same number of groups. This model would be expected to *massively overfit*. Learning experience for all!