

PS 787 Assignment 6 (Dec 7, 2007, due Dec 14)

Consider the 2008 National Annenberg Election Study (NAES). For a detailed description of the planned survey, see See the Annenberg publication *On the Record* (Sept. 2007, vol. 1, number 2.) at

<http://www.appcpenn.org/ShowPage.aspx?myName=ontherecord>

The implementation details of the survey as described here may differ from the actual implementation (to enhance the feasibility of this assignment).

The panel has a wave of interviews with a random sample of adult Americans conducted during October-December 2007, then a second wave reinterviewing the same people during January 1–March 31, 2008. The sample size for the initial wave is large ($n = 18,200$). Attrition is expected to produce a second-wave sample size of $n = 14,560$. Survey participants respond to questionnaires presented to them over the internet (using equipment provided by the survey organization where needed). Within each wave, the precise day each respondent is prompted to respond to a survey is selected at random, with an equal number assigned to each day during each period. Hence during each period there are $18200/91 = 200$ people solicited to respond each day. Random assignment of people to days is independent both within and across waves. It is expected that half of the respondents who eventually complete each survey will do so by the fourth day after they are solicited.

My focus is on estimating the effect that the events of February 5, 2008, have on individuals' reported partisanship. Specifically, assume partisanship is measured using responses to the following question: "Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?" The survey will also include a long list of standard demographic questions (age, sex, education, income, religion, church attendance, marital status, number of children in household, state the respondent lives in), as well as questions about political involvement and opinions (pay attention to politics, to the presidential campaigns; did or will the respondent vote in a primary, in 2004; ideological self-placement, position on various issues, perceptions of the candidates and parties). Assume that the same questions are asked in each wave, perhaps with minor wording changes to suit timing concerns.

February 5 is the day a large number of states will have their presidential primary elections. For the primary schedule, see

http://www.ncsl.org/programs/legismgt/elect/2008_Pres_Primary_Calendar.htm

http://uspolitics.about.com/od/2008elections/a/prez_primary.htm

It is widely expected that the large fields of candidates who are running for the major parties' nominations will be strongly reduced by the voting that day. Voting on February 5 may well decisively choose the winner for either the Democratic party or Republican party. Either the personal characteristics or the policy positions of the nominees may cause some voters to reconsider their partisan commitments.

Describe what you think would be the best way to use the NAES data to estimate the effects the events of February 5 have on individuals' partisanship. Explain clearly how the random assignment of people to days does or does not facilitate estimating the effects. Compare the strengths and weaknesses of the best design using randomization and the Neyman-Rubin model to the best specification using a regression or regression-like approach. Be specific and explicit in describing the various estimators that come into play.