

ST433/533 – Final Exam – Due 11/20

Presidential election polling and subsequent analysis have emerged as one of the most public applications of statistical methods. These spatiotemporal data are noisy and potentially biased. Your objective in this exam is to study the bias in state-level Presidential election polls from the elections of 2012, 2016 and 2020.

The data for this exam are provided on the course webpage, and include polls and final results for each state and election for 2012 and 2016 and links to streaming data for 2020. Define Y_{it} as the percentage of the actual votes in the election of year t in state i that were cast for the Republican (GOP) candidate (you may consider discarding third-party votes), and X_{it} be the corresponding polling average. The polling bias is the expected difference, $B_{it} = E(Y_{it} - X_{it})$. Your objectives are:

1. Devise a method to combine the individual polls to forecast the election results in each state and each year. Your polling averages should be an average of the polls,

$$X_{it} = \sum_{j=1}^{N_t} w_{ijt} P_{jt},$$

where N_t is the total number of polls in election year t , P_{jt} is poll j 's estimated percent GOP support and the weights w_{ijt} sum to one, i.e., $\sum_{j=1}^{N_t} w_{ijt} = 1$, for all i and t . Your method should up-weight polls taken closer to the election and in state i (Nate Silver has spent his career on this step, but you should do something simple and move on to the other objectives).

2. Test whether there is systematic polling bias under the assumption that the bias is constant over state and election.
3. Test whether the bias varies by state and/or election and display the estimated bias. This objective is optional for ST433.

Objectives 2 and 3 should be addressed using the spatial models covered in class. You should also examine sensitivity to the tests to the definition of the polling average in Objective 1.

The final product is a 10-15 minute presentation. Your slides should include sufficient detail that the results could be reproduced by another student in the class. You will present your results to the class over zoom starting at noon on November 20. Email your slides to the instructor (bjreich@ncsu.edu) by midnight on the 19th.

ID	Team
1	Adams, Harris, Peng
2	Bai, Fan, Pena
3	Carbajal Carrasco, Goodman, Hutchens
4	Colonnese, Freedman, Tabor
5	Dixit, Fidan, H Jiang
6	Foraker, Pammer
7	Hasnat, Watson
8	Z Jiang, Schappe, Zheng
9	Rangwala, Turner
10	Schulte, Xu, Wang
11	Wiecha, Yang, Yao