Report 3: ANES

Madison Lewis, Unity Tambellini-Smith, Uma Nagarajan-Swenson, and Grayson Shaw
Politics 90, Spring 2022
Due: May 4, 11:59pm (Wednesday), extension for groups members missing
Where: Upload to Gradescope

Introduction

Does party affiliation affect an individual's likelihood to vote?

Previous scholarship suggests low voter turnout may adversely affect Democratic candidates, and in light of the approaching 2022 midterm elections, it seems prescient to use the newest available data to examine whether this theory holds. Using data from the American National Election Survey in 2020, we use logistic regression to test whether increasing conservative affiliation makes an individual more likely to vote. We find that when other possible explanatory variables are controlled for, party ID does not correlate significantly with decision to vote.

Theory

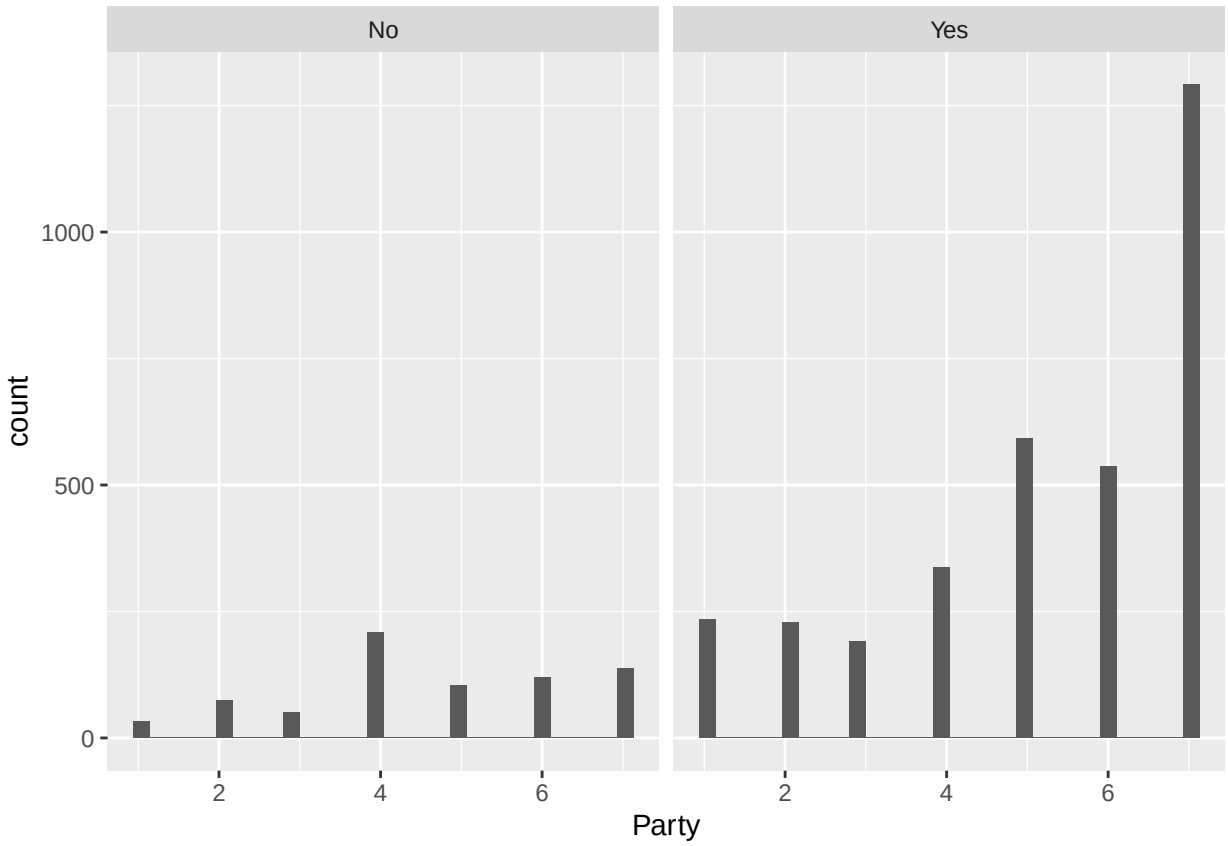
The question of party affiliations effects on voter turnout has been raised in several different studies throughout the past few decades. While some studies (Radcliff 1994) contend that lower voter turnout negatively affects democratic presidential candidates, others find that it does not have an effect either way (Citrin, Shickler, and Sides 2003; Martinez & Gill 2005). Those that find no effect typically posit that other variables, such as class, education, or year undermine the relationship noted in the studies that do find a correlation. Most of these studies examine the relationship, specifically, between turnout and party outcome.

Data

Instead of regressing party outcome on turnout, from the ANES dataset we selected Party ID and A binary measure of whether or not the individual voted in the presidential election as our explanatory and outcome variables, respectively. Party ID was originally on a scale from "strongly democratic" to "strongly republican" with the midpoint at "independent," however, to make it continuous we recoded it to be numeric, on a scale from 1 to 7 with 1 representing most democratic, and 7 representing most republican. For control variables we selected (1) age, (2) education level, (3) sex, (4) health status, (5) satisfaction with life, and (6) whether or not a member of the household had had COVID-19. Some of these control variables (1,2,3,5) were selected because they seemed like they may cause some reasonable variance in a person's likelihood to vote, others (4,6) were selected because they seemed like they may correlate to a person's ability to vote, for example, poor health status may lead individuals to have a harder time voting.

Table 1: Summary Statistics

	Overall (N=4139)
Party ID	
Independent	546 (13.2%)
Independent-Democrat	242 (5.8%)
Independent-Republican	695 (16.8%)
Somewhat Democrat	303 (7.3%)
Somewhat Republican	657 (15.9%)
Very Democrat	267 (6.5%)
Very Republican	1429 (34.5%)
Sex	
N-Miss	22
Female	2106 (51.2%)
Male	2011 (48.8%)
Age	
N-Miss	3988
Age 18-34	28 (18.5%)
Age 35-50	44 (29.1%)
Age 51-64	64 (42.4%)
Age 65+	15 (9.9%)
Did you vote in 2020?	
No	729 (17.6%)
Yes	3410 (82.4%)
Health Status	
N-Miss	35
Excellent	588 (14.3%)
Fair	606 (14.8%)
Good	1404 (34.2%)
Poor	154 (3.8%)
Very Good	1352 (32.9%)
Positive Household COVID-19 Test?	
N-Miss	28
No	3958 (96.3%)
Yes	153 (3.7%)
Satisfied with Life?	
N-Miss	44
No	405 (9.9%)
Yes	3690 (90.1%)

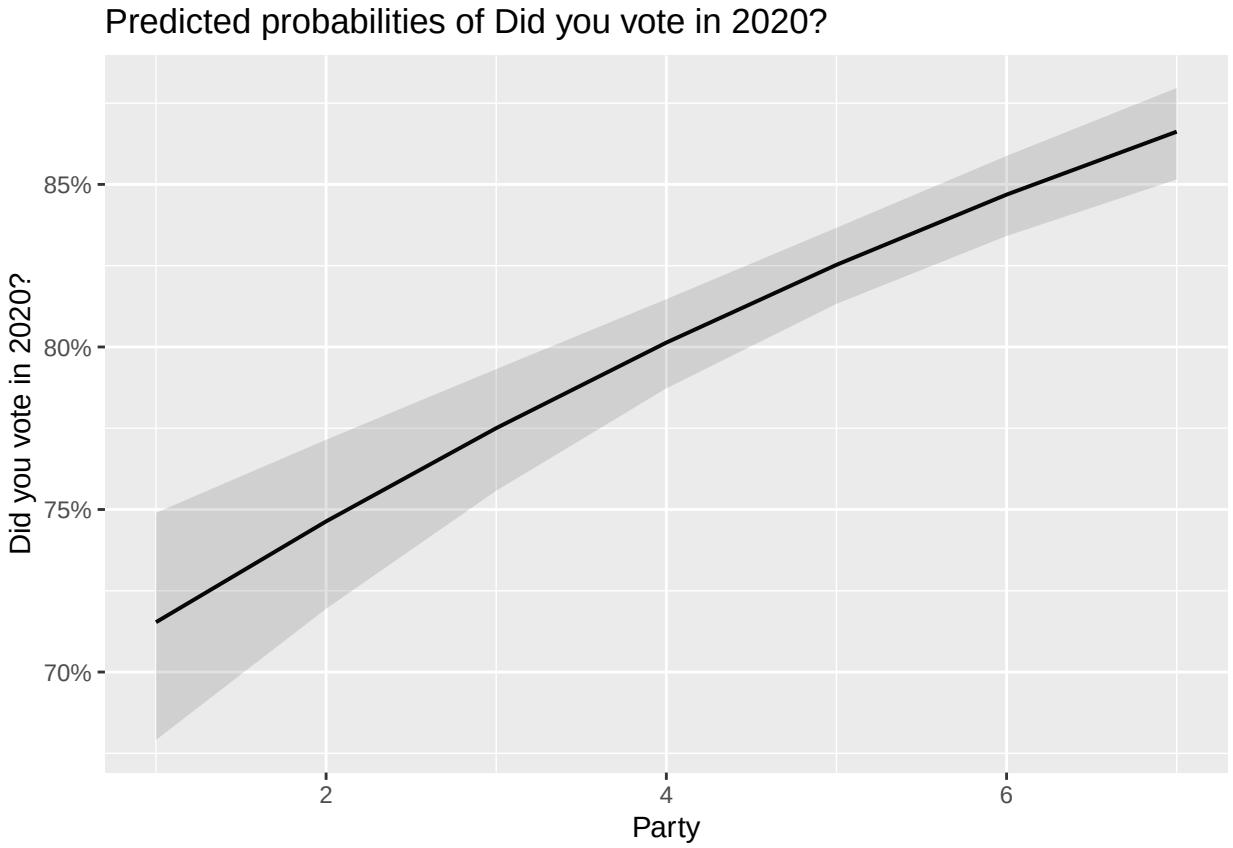


Methods

We used the statistical test glm(), a linear regression test specifically designed for logistic regression, which was appropriate given that our outcome variable is binary. Logistic regression broadly requires that the data has a homogeneity of variance, normal residuals, and influential observations. Our tests (Appendix A) showed that there may be violations to the first two assumptions, as the homogeneity of the variance line was horizontal but not flat and some residuals did not fall along the desired line. For our logistic regression we ran three models with increasingly more variables controlled for. The first model looked only at the relationship between party ID and voter turnout, the second at voter turnout on party ID with age, education, and sex controlled for, and the third the same as the second but with health status, satisfaction, and COVID status also controlled for.

Results

```
## $Party
```



$$\log\left[\frac{P(\text{Did you vote in 2020? = Yes})}{1-P(\text{Did you vote in 2020? = Yes})}\right] = \alpha + \beta_1(\text{Party}) \tag{1}$$

$$\log\left[\frac{P(\text{Did you vote in 2020? = No})}{1-P(\text{Did you vote in 2020? = No})}\right] = \alpha+\beta_1(\text{Party})+\beta_2(\text{Age }_{35\text{-}50})+\beta_3(\text{Age }_{51\text{-}64})+\beta_4(\text{Age }_{65+})+\beta_5(\text{`Education Level`} \tag{2}$$

$$\log\left[\frac{P(\text{Did you vote in 2020? = No})}{1-P(\text{Did you vote in 2020? = No})}\right] = \alpha+\beta_1(\text{Party})+\beta_2(\text{Age }_{35\text{-}50})+\beta_3(\text{Age }_{51\text{-}64})+\beta_4(\text{Age }_{65+})+\beta_5(\text{`Education Level`} \tag{3}$$

The first plot shows that there is a significant positive correlation between party and likelihood to vote, so at first the naive model looks appealing. However, the stargazer table shows how the significance of party ID reduces when other factors are introduced. This would indicate that other factors contribute more to voter turnout, such as education, compared to party ID.

The naive model found a strong correlation (p value below .005) between party ID and voter turnout. However, our second and third models failed to find a significant correlation when other variables were controlled for, suggesting other factors (for example, education level) may provide a stronger explanation for why individuals do or do not vote. An ANOVA test (Appendix B) suggested that there is not a significant advantage for using one model over another with a p-value of .13 and .14 between the second and third model. This means that the models are equally useful for showing the data, but the confounding variables in the intermediate and complex model show that there are other factors beyond party ID. Therefore, we fail to reject the null hypothesis, and do not find that party ID has a significant effect on voter turnout.

Table 2:

	Dependent variable:		
	‘Did you vote in 2020¿		
	(1)	(2)	(3)
Party	0.16*** (0.02)	0.15 (0.19)	0.15 (0.21)
AgeAge 35-50		0.48 (1.26)	−0.20 (1.48)
AgeAge 51-64		0.86 (1.07)	0.08 (1.29)
AgeAge 65+		0.06 (1.41)	17.60 (4,974.00)
‘Education Level‘Bachelors		0.63 (1.55)	1.15 (1.64)
‘Education Level‘Doctorate		16.20 (5,194.00)	18.30 (8,706.00)
‘Education Level‘Graduate School		16.90 (3,253.00)	18.60 (5,559.00)
‘Education Level‘High School		−1.28 (1.24)	0.17 (1.36)
‘Education Level‘Less than high school		−4.34** (1.99)	−2.65 (2.36)
SexMale		−1.21 (0.94)	−0.29 (1.06)
‘Health Status‘Fair			−0.76 (1.66)
‘Health Status‘Good			0.43 (1.59)
‘Health Status‘Poor			−56.60 (14,002.00)
‘Health Status‘Very Good			0.05 (1.39)
‘Satisfied with Life¿Yes			−17.70 (10,347.00)
‘Positive Household COVID-19 Test¿Yes			18.40 (17,730.00)
Constant	0.76*** (0.11)	2.36 (1.80)	19.10 (10,347.00)
Observations	4,139	88	88
Log Likelihood	−1,898.00	−22.20	−17.40
Akaike Inf. Crit.	3,800.00	66.30	68.80
Note:			*p<0.1; **p<0.05; ***p<0.01

Discussion

Our model, unlike other models, is focused on whether party ID affects voter turnout, rather than the more popular method of looking at if low voter turnout affects outcome (ie. if years with low voter turnout are correlated with Republican victors), as has been utilized in most other studies. Because of this our data is limited only to one year, and does not directly examine the outcomes of given races, so it is possible that the scope of inference is limited and not applicable to the total population. This, in turn, limits our results because the true causative factor behind the election results may be a specific condition of that year alone as opposed to a more established condition like party ID. Furthermore, as the ANES data is observational, it would be difficult to draw causal inferences without employing further techniques, matching, for example.

Conclusion

Through linear regression, we did not find evidence that party ID affects voter turnout in agreement with Citrin and Martinez and in disagreement with Radcliff. The lack of sustained significance in the data indicates that there is more to the question of voter turnout than a single factor, ultimately indicating that there are a variety of contributors for why an individual decides to (or decides not to) vote.

Bibliography

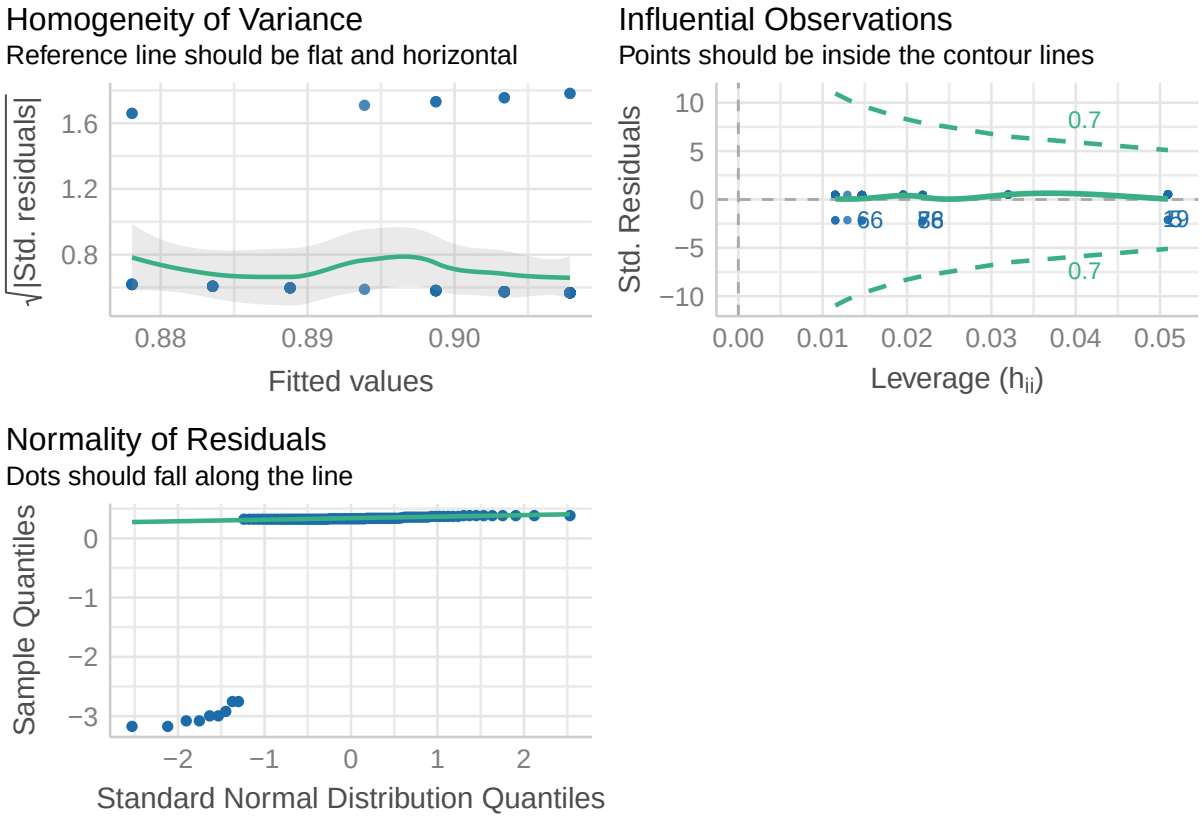
Citrin, Jack, Eric Schickler, and John Sides. 2003. "What If Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections." American Journal of Political Science 47(1): 75–90.

Radcliff, Benjamin. 1994. "Turnout and the Democratic Vote." American Politics Quarterly 22: 259–76.

Martinez, Michael D, and Jeff Gill. "The Effects of Turnout on Partisan Outcomes in U.s. Presidential Elections 1960-2000." The Journal of Politics 67, no. 4 (2005): 1248–74.

Appendix A

Performance Test



Appendix B

ANOVA results

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        86       58.0 NA       NA       NA
## 2        77       44.3  9     13.7    0.134
## 3        71       34.8  6      9.5    0.147
```