

Turn Out Data: In-Class Exercise

Surveys are frequently used to measure political behavior such as voter turnout, but some researchers are concerned about the accuracy of self-reports. In particular, they worry about possible social desirability bias where, in postelection surveys, respondents who did not vote in an election lie about not having voted because they may feel that they should have voted. Is such a bias present in the American National Election Studies (ANES)? ANES is a nationwide survey that has been conducted for every election since 1948. ANES is based on face-to-face interviews with a nationally representative sample of adults. Table 1 displays the names and descriptions of the variables in the turnout.csv data file.

Table 1: US Election Turnout Data Codebook

Variable	Description
ANES	ANES's estimated voter turnout rate
VEP	Voting eligible population (in thousands)
VAP	Voting age population (in thousands)
total	Total ballots cast in the election (in thousands)
felons	Total number of ineligible voters due to their felon status (in thousands)
noncit	Total number of ineligible voters due to their noncitizen status (in thousands)
overseas	Total eligible overseas voters (in thousands)
psvoters	Total ballots cast by overseas voters (in thousands)

Part 1

1. Generate a new, blank R Script and save it in your R folder. Name it: lastname_turnout. So, for example, mine would be called: Porter_turnout.R
2. Start by identifying yourself, writing the date, and describing this script. Then, clear out your workspace.
3. Set your working directory.
4. Load the turnout dataset into R and assign it to an object called *turnout*. Then, take a look at the data using the View() function.
5. What are the dimensions (i.e. number of observations and number of variables) of the dataset? Remember: report your answer in the format: rows x columns.
6. Look at a summary of all of the variables in the data. What is the range of years covered in the dataset?

Part 2

Now, we are going to start to deal with turnout rates. A general voter turnout rate would be calculated like:

$$\left(\frac{\textit{Total number of people who voted}}{\textit{Total number of people who COULD have voted}} \right) * 100$$

We know the total number of people who voted – that never changes. BUT we can obtain different rates for voter turnout depending on how we measure the total number of people who COULD have voted. Sometimes, this is measured with the VAP (voting age population), which is simply the total number of people whose age qualifies them to vote. This measure is not perfect, though, because people can be disqualified from voting for reasons beyond age, such as citizenship status or felony record (side note: restrictions based on felony record varies by state). These questions deal with the differences when you calculate voter turnout rates based on voting age population (VAP) or voting eligible population (VEP).

7. Calculate the turnout rate using the voting eligible population (VEP). Use the formula above to guide you. Call this new vector *VEPtr*. This gives you the turnout rate for each year, by voting eligible population.
8. Now, calculate the turnout rate based on the voting age population (VAP) for each year and assign this turnout rate to an object called *VAPtr*. Note for this calculation, we must add the total number of eligible overseas voters to the VAP (in the denominator of the equation) since the VAP variable does not include these individuals in the count (but, the total variable does include overseas ballots (osvoters) so no need to modify that). This gives you the turnout rate for each year, by voting age population.
9. Bind these two vectors to the turnout data frame.

Part 3

Now you have the voter turnout rate based on voting age population (VAP) and voting eligible population (VEP). You're interested in how these rates compare to self-reported voter turnout rates. Do people lie about whether or not they voted? That is, is there social desirability bias present in the self-reported voter turnout rate in the ANES? These questions ask you to compare the voter turnout rates you calculated (based on VAP and VEP) to the ANES's turnout rate. Note that this dataset includes a variable that *is* ANES turnout rate (called "ANES") so you don't have to calculate this with the same equation you used in Part 2.

10. Compute the **difference** between the VAP turnout rate (that you calculated) and ANES estimates of turnout rate (the variable: ANES). What is the average difference between the VAP and ANES estimated turnout rates? What is the range of the difference?

11. Now, conduct the same comparison between VEP turnout rate and the ANES estimates of turnout rate and assign the difference to an object called *diffVEP*. What is the average difference there? What is the range of differences?
12. Which estimate – the VAP or VEP – of voter turnout is closer to the ANES estimate, on average? Why do you think that is? Is it starting to seem like there may be social desirability bias present in the ANES reports of voter turnout?

Part 4

You start to think that there may be certain times when people are more likely to lie about voting than others. Voter turnout rates are lower in midterm elections – so maybe people are more willing to admit that they don't vote during the midterms. Are people more likely to lie about their participation in presidential elections than in midterm elections? That is what you will investigate here!

13. Compare the VEP turnout rate with the ANES turnout rate separately for presidential elections and midterm elections. Make sure to note that the dataset excludes the year 2006.
14. Does the bias of the ANES estimates vary across election type? Summarize your two new vectors and report what you see.

Part 5

After all that work, you realize that the ANES does not interview prisoners and overseas voters. So, you decide to calculate an adjustment to the VAP turnout rate – but only for 2008. Follow the steps below to calculate an adjustment to the VAP turnout rate.

15. Begin by subtracting the total number of ineligible felons and noncitizens from the VAP to calculate an adjusted VAP. Create a new **variable** called *adj_VAP*.
16. Next, calculate an adjusted VAP turnout rate. Here, take care to subtract the number of overseas ballots counted from the total ballots in 2008. Make sure to use your newly created variable, *adj_VAP* in the denominator. Read this rate into an object called *adj_VAPtr*