

Homework 1

YOUR NAME HERE

SUBMISSION DATE HERE

0.1

Complete Swirl lesson INTRO2.

Briefly describe one technique or concept in R that you are having trouble understanding. And don't worry! It is confusing and hard!

Bias in voter turnout (book exercise 1.5.1)

You can obtain the file needed for this assignment either from the `qss.student` package, or from my web page:

https://raw.githubusercontent.com/f-edwards/intro_stats/master/hw/data/turnout.csv

Save the file as `turnout.csv` in the working directory that your homework file is in.

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 post-election 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)? The ANES is a nation-wide survey that has been conducted for every election since 1948. The ANES conducts face-to-face interviews with a nationally representative sample of adults. The table below displays the names and descriptions of variables in the `turnout.csv` data file.

Name	Description
<code>year</code>	Election year
<code>VEP</code>	Voting Eligible Population (in thousands)
<code>VAP</code>	Voting Age Population (in thousands)
<code>total</code>	Total ballots cast for highest office (in thousands)
<code>felons</code>	Total ineligible felons (in thousands)
<code>noncitizens</code>	Total non-citizens (in thousands)
<code>overseas</code>	Total eligible overseas voters (in thousands)
<code>osvoters</code>	Total ballots counted by overseas voters (in thousands)

Question 1

Load the data into R and check the dimensions of the data. Also, obtain a summary of the data. How many observations are there? What is the range of years covered in this data set?

Question 2

Calculate the turnout rate based on the voting age population or VAP. Note that for this data set, we must add the total number of eligible overseas voters since the `VAP` variable does not include these individuals in the count. Next, calculate the turnout rate using the voting eligible population or VEP. What difference do you observe?

Question 3

Compute the difference between VAP and ANES estimates of turnout rate. How big is the difference on average? What is the range of the difference? Conduct the same comparison for the VEP and ANES estimates of voter turnout. Briefly comment on the results.

Question 4

Compare the VEP turnout rate with the ANES turnout rate separately for presidential elections and midterm elections. Note that the data set excludes the year 2006. Does the bias of the ANES vary across election types?

Question 5

Divide the data into half by election years such that you subset the data into two periods. Calculate the difference between the VEP turnout rate and the ANES turnout rate separately for each year within each period. Has the bias of the ANES increased over time?

Question 6

The ANES does not interview overseas voters and prisoners. Calculate an adjustment to the 2008 VAP turnout rate. Begin by subtracting the total number of ineligible felons and non-citizens from the VAP to calculate an adjusted VAP. Next, calculate an adjusted VAP turnout rate, taking care to subtract the number of overseas ballots counted from the total ballots in 2008. Compare the adjusted VAP turnout with the unadjusted VAP, VEP, and the ANES turnout rate. Briefly discuss the results.