Lab 1: R building blocks

Bias in self-reported turnout

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)? 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 variables that you will create in your environment if you copy, paste, and run the following code in R.

```
year <- c(1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998,
          2000, 2002, 2004, 2008)
VEP <- c(159635, 160467, 167702, 170396, 173579, 176629, 179656, 182623,
         186347, 190420, 194331, 198382, 203483, 213314)
VAP <- c(164445, 166028, 173995, 177922, 181955, 186159, 190778, 195258,
         200016, 205313, 210623, 215462, 220336, 230872)
total <- c(86515, 67616, 92653, 64991, 91595, 67859, 104405, 75106, 96263,
           72537, 105375, 78382, 122295, 131304)
ANES <- c(71, 60, 74, 53, 70, 47, 75, 56, 73, 52, 73, 62, 77, 78)
felons <- c(802, 960, 1165, 1367, 1594, 1901, 2183, 2441, 2586, 2920, 3083,
            3168, 3158, 3145)
noncitizens <- c(5756, 6641, 7482, 8362, 9280, 10239, 11447, 12497, 13601,
                 14988, 16218, 17237, 18068, 19392)
overseas <- c(1803, 1982, 2361, 2216, 2257, 2659, 2418, 2229, 2499, 2937,
              2937, 3308, 3862, 4972)
osvoters <- c(NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 263)
```

Table 1: US Election Turnout Data.

Variable	Description
`year`	election year
`ANES`	ANES estimated turnout rate (as percentage)
`VEP`	voting eligible population (in thousands)
`VAP`	voting age population (in thousands), excluding overseas
`total`	total ballots cast for highest office (in thousands)
`felons`	total ineligible felons (in thousands)
`noncitizens`	total noncitizens (in thousands)
`overseas`	total eligible overseas voters (in thousands)
`osvoters`	total ballots counted by overseas voters (in thousands)

Question 1

Create the variables in R and check the dimensions of the the variables. Also, obtain a summary of the data (min and max, mean, standard deviation). How many observations are there? What is the range of years covered in this data set?

Hints:

- `summary`, `length`, `mean`, `sd`, `range`, `min`, `max`
- `lapply( list(a, b, c), length )` is equivalent to `list( length(a) , length(b), length(c) )`
- `sapply( list(a, b, c), length )` is equivalent to `c( length(a) , length(b), length(c) )`

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?

Hints:

- Name the entries of the turnout rates with the year. To do so, you can use the `names()` function:

```
v <- c(20, 50)
names(v) <- c("AAAA", "BBBB")
```

```
v
```

```
AAAA BBBB
  20   50
```

Question 3

Compute the differences between the VAP and ANES estimates of turnout rate. How big is the difference on average? What is the range of the differences? 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 estimates vary across election types?

Hints:

- You can use either numeric or logical indexing.
- Presidential elections are in years that are multiples of 4 see this Wikipedia article.

  To check if a is a multiple of 4, you can check if the remainder of the division by 4 is 0: `a %% 4 == 0`.
- Midterm elections are in years that are not multiples of 4, see this Wikipedia article.

Question 5

Imagine dividing the years into two periods of seven years each. Calculate the average difference between the VEP turnout rate and the ANES turnout rate separately for each period. Has the bias of ANES increased over time?

Hints:

- You can use either numeric or logical indexing.

Question 6

ANES does not interview prisoners and overseas voters. Calculate an adjustment to the 2008 VAP turnout rate. Begin by subtracting the total number of ineligible felons and noncitizens 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 summarise the results.

Source

The exercises have been taken from Chapter 1 of

- Imai, K. (2018). Quantitative social science: an introduction. Princeton University Press.