Estimating the Bias in Self-Reported Turnout
Part II: Computing and Interpreting Means

Let's continue working with the official and the self-reported ANES turnout data from 1980 to 2004. The dataset we will use is in a file called "ANES.csv". Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is federal elections in the U.S.

variable	description
year	year of the election
presidential	whether it was a presidential election: 1=yes, 0=no
midterm	whether it was a midterm election: 1=yes, 0=no
ANES_turnout	proportion of ANES respondents who reported to have voted in the election (in percentages)
votes	number of ballots officially cast in the election (in thousands)
VEP	voting eligible population at the time (in thousands)
VAP	voting age population at the time (in thousands)
felons	number of felons not eligible to vote (in thousands)
noncitizens	number of non-citizens living in the U.S. (in thousands)

Table 1: Variables in "ANES.csv"

In this problem set, we practice (i) using arithmetic operators to create new variables as well as (ii) computing and interpreting means.

As always, we start by loading and looking at the data:

```
## load and look at the data
anes <- read.csv("ANES.csv") # reads and stores data
head(anes) # shows first observations
##   year presidential midterm ANES_turnout votes   VEP   VAP felons noncitizens
## 1 1980            1       0           71 86515 159635 164445    802        5756
## 2 1982            0       1           60 67616 160467 166028    960        6641
## 3 1984            1       0           74 92653 167702 173995   1165        7482
## 4 1986            0       1           53 64991 170396 177922   1367        8362
## 5 1988            1       0           70 91595 173579 181955   1594        9280
## 6 1990            0       1           47 67859 176629 186159   1901       10239
```

1. Create a new variable called *VEP_turnout* defined as the number of ballots officially cast in the election divided by the voting eligible population and multiplied by 100. Make sure to store this new variable in the existing dataframe named *anes* by using the character $. (See page 41 of DSS, to learn how to use the character $ to identify a variable inside a dataframe not just to access it but also to create it.) (5 points)
2. Use the function head() to look at the first few observations again to ensure that you have created the new variable, *VEP_turnout*, correctly. Is the first value of *VEP_turnout* what one

would expect, given the first values of *votes* and *VEP*? What is the unit of measurement of *VEP_turnout*? (5 points)

3. Now, create a new variable called *VAP_turnout* defined as the number of ballots officially cast in the election divided by the voting age population and multiplied by 100. Make sure to store this new variable in the existing dataframe named *anes*. (5 points)
4. Use the function head() to look at the first few observations again to ensure that you have created the new variable, *VAP_turnout*, correctly. Is the first value of *VAP_turnout* what one would expect, given the first values of *votes* and *VAP*? What is the unit of measurement of *VAP_turnout*? (5 points)
5. Looking at the first few observations of the two new variables, *VEP_turnout* and *VAP_turnout*, shown by the function head() above, can you tell whether one of them always contains higher values than the other? Why do you think that is? Which of the two variables do you think most accurately measures turnout? (5 points)
6. Use the function mean(), to compute the average value of *VEP_turnout* among the 13 federal elections in the dataset. Please provide a full substantive interpretation of what this average means and make sure to provide the unit of measurement. (10 points)
7. Use the function mean(), to compute the average value of *ANES_turnout* among the 13 federal elections in the dataset. Please provide a full substantive interpretation of what this average means and make sure to provide the unit of measurement. (10 points)
8. When comparing the average value of *ANES_turnout* to the average value of *VEP_turnout*, do you find any evidence of people lying about their voting behavior? (5 points)