What is the Effect of the Death of the Leader on the Level of Democracy? Part II: Fitting a Line to Compute the Difference-in-Means Estimator

(Based on Benjamin F. Jones and Benjamin A. Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics*, 1 (2): 55-87.)

There is a longstanding debate in the study of international relations on whether individual political leaders make a difference. To explore this issue, let's estimate the causal effect of the death of the leader on the level of democracy of a country. For this purpose, we will analyze data on assassination attempts against political leaders from 1875 to 2004.

To measure the level of democracy of the country, we will use polity scores. Polity scores categorize the regime of a country on a 21-point scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy). The Polity Project has produced polity scores for all countries from 1800 and on. For example, here are the 2018 polity scores.

The dataset is in a file called "leaders.csv". Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is assassination attempts.

variable	description
year	year of the assassination attempt
country	name of the country where the assassination attempt took place
leadername	name of the leader whose life was at risk in the assassination attempt
died	whether the leader died as a result of the assassination attempt: 1=yes, 0=no
politybefore	polity scores of the country where the assassination attempt took place before the assassination attempt (in points, in a scale from -10 to 10)
polityafter	polity scores of the country where the assassination attempt took place after the assassination attempt (in points, in a scale from -10 to 10)

Table 1: Variables in "leaders.csv"

Whether an assassination attempt occurs or not is not a random process. (For example, there are probably more assassination attempts in dictatorships than in full democracies.) However, once an assassination attempt has occurred, one could argue that whether the assassination attempt is successful or not is the result of small elements of randomness, such as the timing and path of the weapon. As a result, we can consider (at least for now) that, after an assassination attempt, the death a leader is close to random and, thus, the assassination attempts where the leader ended up dying should be, on average, comparable to the assassination attempts where the leader ended up surviving. If this is true, then we can estimate the average causal effect of the death of the leader by computing the difference-in-means estimator.

In this problem set, we practice fitting a linear model to compute the difference-in-means estimator.

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

```
## load and look at the data
leaders <- read.csv("leaders.csv") # reads and stores data
head(leaders) # shows first observations
##   year      country       leadername died politybefore polityafter
## 1 1929 Afghanistan Habibullah Ghazi    0           -6   -6.000000
## 2 1933 Afghanistan       Nadir Shah    1           -6   -7.333333
## 3 1934 Afghanistan      Hashim Khan    0           -6   -8.000000
## 4 1924      Albania             Zogu    0            0   -9.000000
## 5 1931      Albania             Zogu    0           -9   -9.000000
## 6 1968      Algeria       Boumedienne    0           -9   -9.000000
```

1. First, let's identify our Y and X variables. Given that we are interested in estimating the average causal effect of the death of a leader on the polity scores of a country:
    a. What should be our Y variable? In other words, which variable is the outcome variable? And, is this variable binary or non-binary? (5 points)
    b. What should be our X variable? In other words, which variable is the treatment variable? And, is this variable binary or non-binary? (5 points)
2. Compute the difference-in-means estimator directly and report its value. (5 points)
3. Now, let's use the `lm()` function to fit a line to the data and summarize the relationship between X and Y. (Hint: The `lm()` function requires an argument of the form Y~X) (R code only) (5 points)
4. What is the fitted line? In other words, provide the formula $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X$ where you specify each term (i.e., substitute Y for the name of the outcome variable, substitute $\widehat{\alpha}$ for the estimated value of the intercept coefficient, substitute $\widehat{\beta}$ for the estimated value of the slope coefficient, and substitute X for the name of the treatment variable.) (5 points)
5. Is the estimated slope coefficient ($\widehat{\beta}$) equivalent to the value of the difference-in-means estimator in this case? A yes or no answer will suffice. (5 points)
6. Please provide a full substantive interpretation of the estimated slope coefficient (including the unit of measurement). (10 points)
7. What is the average causal effect of the death of a leader on the polity scores of a country? (Please write a full sentence answering the question, including the assumption, why the assumption might be reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (10 points)