# What is the Effect of the Death of the Leader on the Level of Democracy? Part III: Controlling for Confounders (with Solutions)

Let's continue to work with the data on assassinations attempts against political leaders from 1875 to 2004. 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"

In this problem set, we will revisit the analysis we performed in *What is the Effect of the Death of the Leader on the Level of Democracy? Part II* in view of what we learned in *Part I*.

In this problem set, we practice estimating causal effects with observational data by fitting a multiple linear regression model where $Y$ is the outcome variable of interest, $X_1$ is the treatment variable, and $X_2$ is the confounding variable we worry about.

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. Let's start by quickly replicating the results from Part II: Use the lm() function to fit a line to the data and summarize the relationship between X and Y and answer the question: 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) (5 points)

   R code:

   ```
   lm(leaders$ polityafter ~ leaders$died) # or
   ```

   ```
   lm( polityafter ~ died, data = leaders)
   ##
   ## Call:
   ## lm(formula = polityafter ~ died, data = leaders)
   ##
   ## Coefficients:
   ## (Intercept)         died
   ##      -1.895        1.132
   ```

   Answer: Assuming that the assassination attempts where the leader ended up dying are comparable to the assassination attempts where the leader ended up surviving (an assumption that might be reasonable if the death of the leader after an assassination attempt is close to random), we estimate that the death of the leader increases the country's polity scores after the assassination attempt by 1.13 points, on average.

2. Note that, as stated in our conclusion statement above, the validity of the estimate we arrived at in Part II depends on whether the assumption we made is correct (that is, on whether the assassination attempts where the leader ended up dying are comparable to the assassination attempts where the leader ended up surviving). Given your answers to questions 4 and 5 in Part I, do you think the assumption is correct? (5 points)

   Answers: No, the assumption is not correct. As we concluded in question 5 of Part I, a smaller proportion of successful assassination attempts have a value of *politybefore* of -10. And, as we concluded in question 5 of Part I, countries where the assassination attempt ended up being successful were, on average, slightly more democratic to begin with (that is, before the assassination took place) than countries where the assassination attempt ended up not being successful. (Note: The average level of democracy before the assassination attempts is less negative for successful assassination attempts than for unsuccessful assassination attempts.)

3. Given what you learned in Part I about (1) the relationship between *politybefore* and *died* (questions 4 and 5) and (2) the relationship between *politybefore* and *polityafter* (questions 6 and 7), do you think *politybefore* is a confounder? Please provide your reasoning. (10 points)

   Answers: Yes, *politybefore* appears to be a confounding variable since: (1) as we learned in questions 4 and 5 in part 1, *politybefore* seems to affect the likelihood of the leader dying (that is, it affects the likelihood of receiving the treatment), and (2) as we learned in questions 6 and 7, *politybefore* and *polityafter* are moderately to highly correlated with each other, in other words, *politybefore* seems to affect the value of the level of democracy after the assassination attempt (that is, it affects our outcome variable of interest). (Note that for *politybefore* to be a confounding variable it must affect both the value of *died* and the value of *polityafter*.)

$$\begin{tikzcd} & \textit{politybefore} \arrow[dl] \arrow[dr] & \\ \textit{died} \arrow[rr] & & \textit{polityafter} \end{tikzcd}$$

4. Now, let's estimate the average causal effect of *died* on *polityafter*, while controlling for *politybefore*. Start by fitting the following linear model with R:

$$\widehat{polityafter} = \widehat{\alpha} + \widehat{\beta}_1 \; died + \widehat{\beta}_2 \; politybefore$$

(Hint: To fit a multiple linear regression model in R, we can use the function lm() and specify as the main required argument a formula of the type $Y \sim X_1 + X_2 + ... + X_p$.)

What is the fitted model? (5 points)

R code:

```
lm(leaders$polityafter ~ leaders$died + leaders$politybefore) # or
```

```
lm(polityafter ~ died + politybefore, data = leaders)
##
## Call:
## lm(formula = polityafter ~ died + politybefore, data = leaders)
##
## Coefficients:
## (Intercept)         died   politybefore
##     -0.4346       0.2616         0.8375
```

Answer: The fitted model is: $\widehat{\text{polityafter}}$ = -0.43 + 0.26 *died* + 0.84 *politybefore*.

5. Assuming *politybefore* is the only confounding variable present, please provide a full substantive interpretation of $\widehat{\beta}_2$ (including the unit of measurement). (10 points)

Answer: When we hold constant the level of democracy of the country before the assassination attempt took place, we estimate that the death of the leader increases the country's polity scores after the assassination attempt by 0.26 points, on average. (Note: We use causal language (i.e., increases), and not just predictive (i.e., is associated with a predicted increase), because we assume that *politybefore* is the only confounding variable present. Since we control for *politybefore* in the estimation process by adding the variable in the linear model as a control, the new estimate of the average treatment effect ($\widehat{\beta}_2$) should be valid if *politybefore* is indeed the only confounder we need to worry about.)

6. Assuming *politybefore* is the only confounding variable present, 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)

   Answer: Let's start by figuring out each key element separately.

   - *What's the assumption?* We assume that the assassination attempts where the leader ended up dying (the treatment group) are comparable to the assassination attempts where the leader ended up surviving (the control group) once we control for the level of democracy of the countries before the assassination attempt took place. (In other words, we assume *politybefore* is the only confounding variable present.)

   - *Why might this assumption be reasonable?* Even though the dataset does not come from a randomized experiment, whether a leader dies after an assassination attempt is arguably close to random once we control for the level of democracy of the countries before the assassination attempt took place.

   - *What's the treatment?* The death of the leader.

   - *What's the outcome?* Polity scores after the assassination attempt.

   - *What's the direction, size, and unit of measurement of the average causal effect?* An increase of 0.26 points, on average. (Note: It is an increase because we are measuring change—the change in the outcome variable caused by the treatment—and the difference-in-means estimator is positive. The unit of measurement is the same as the unit of measurement of $\triangle\overline{Y}$. Since Y is non-binary and measured in points, $\triangle\overline{Y}$, the difference-in-means estimator, and $\widehat{\beta}$ are also measured in points.)

   *Full answer*: Assuming that the assassination attempts where the leader ended up dying are comparable to the assassination attempts where the leader ended up surviving, once we control for the level of democracy of the countries before the assassination attempt took place (an assumption that might be reasonable if the death of the leader after an assassination attempt is close to random, once we control for the level of democracy of the countries before the assassination attempt), we estimate that the death of the leader increases the country's polity scores after the assassination attempt by 0.26 points, on average.

7. Does the size of the average treatment effect change significantly as a result of controlling for the confounder *politybefore*? A yes/no answer will suffice. (5 points)

   Answer: Yes, the size of the average treatment effect becomes much smaller as a result of controlling for the confounder *politybefore*. It goes from being 1.13 to 0.26. (Note: This exemplifies the importance of controlling for confounders! Otherwise, we arrive to conclusions that are completely invalid.)