This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

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 (with Solutions)

(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.

This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

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)

   Answer: The Y variable should be *polityafter*, which is a non-binary variable since it can take more than two values. (Note that since we are trying to measure the effect of the death of the leader on the level of democracy *after* the assassination attempt, *polityafter* should be the outcome variable (Y). The variable *politybefore* should not be affected by the death of the leader. An even better Y variable would be the difference between *polityafter* and *politybefore* so that we can measure the effect that the death of the leader had on the *change* in the level of democracy. However, this is more advanced that what we do in this class.)

   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)

   Answer: The X variable should be *died*, which is a binary variable since it can only take 1s and 0s. (Note: Since we are trying to measure the effect of the death of the leader on the level of democracy, *died* should be the treatment variable (X). Recall, all treatment variables in this class are binary. They equal 1 when the observation was treated, 0 when the observation was not treated. In this case, the treatment is the death of the leader and, thus, the treatment variable is *died*, which equals 1 when the leader died, 0 when the leader did not die.)

2. Compute the difference-in-means estimator directly and report its value. (5 points)

   R code:

```
## compute the difference in means estimator
mean(leaders$ polityafter [ leaders$died==1]) - #average outcome for treatment group
  mean(leaders$ polityafter [ leaders$died==0]) #average outcome for control group
## [1] 1.132212
```

   Answer: The difference-in-means estimator equals 1.13.

This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

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)

   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
```

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)

   Answer: $\widehat{\text{polityafter}}$ = -1.90 + 1.13 *died*. (Note: The Y variable is *polityafter*, $\widehat{\alpha}$=-1.90, $\widehat{\beta}$=1.13, and the X variable is *died*.)

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)

   Answer: Yes, the estimated slope coefficient is equivalent to the difference-in-means estimator. (Note: They are both equal to 1.13.)

6. Please provide a full substantive interpretation of the estimated slope coefficient (including the unit of measurement). (10 points)

   Answer: We estimate that the death of the leader increases the country's polity scores after the assassination attempt by 1.13 points, on average. (Note: The mathematical definition of $\widehat{\beta}$ is the $\triangle\widehat{Y}$ associated with $\triangle X$=1. In this case, $\widehat{\beta}$ is the $\triangle\widehat{\textit{polityafter}}$ associated with $\triangle$*died* = 1. Hence, the death of the leader (that is, when *died* increases by one unit, from 0 to 1) is associated with an increase in polity scores after the assassination attempt of 1.13 points, on average. The unit of measurement is points because Y is non-binary and measured in points so $\triangle\overline{Y}$ should also be in points. In addition, because here X is the treatment variable, $\widehat{\beta}$ is equivalent to the difference-in-means estimator so we should use causal language in its interpretation. Instead of "associated with an increase," we should say "increases" or "causes an increase of." Final answer: We estimate that the death of the leader *increases* the country's polity scores after the assassination attempt by 1.13 points, on average. This average treatment effect should be valid if the assassination attempts where the leader ended up dying are comparable to the assassination attempts where the leader ended up surviving, that is, if there are no confounding variables present.)

This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: *A Friendly and Practical Introduction*. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

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)

   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). (Note: If this assumption were not true the difference-in-means estimator would NOT produce a valid estimate of the average treatment effect.)

   - *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.

   - *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 1.13 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 (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.