Do Women Promote Different Policies than Men? Part III: Estimating an Average Causal Effect (with Solutions)

Let's continue working with the data from the experiment in India. As a reminder, Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is villages.

variable	description
village	village identifier ("Gram Panchayat number _ village number")
female	whether village was assigned a female politician: 1=yes, 0=no
water	number of new (or repaired) drinking water facilities in the village since random assignment
irrigation	number of new (or repaired) irrigation facilities in the village since random assignment

Table 1: Variables in "india.csv"

In this problem set, we practice (1) how to estimate an average treatment effect using data from a randomized experiment and (2) how to write a conclusion statement.

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

```
## load and look at the data
india <- read.csv("india.csv") # reads and stores data
head(india) # shows first observations
##          village female water irrigation
## 1 GP1_village2      1    10          0
## 2 GP1_village1      1     0          5
## 3 GP2_village2      1     2          2
## 4 GP2_village1      1    31          4
## 5 GP3_village2      0     0          0
## 6 GP3_village1      0     0          0
```

1. Considering that the dataset we are analyzing comes from a randomized experiment, what can we compute to estimate the average causal effect of having a female politician on the number of new (or repaired) drinking water facilities? Please provide the name of the estimator. (5 points)

   Answer: The difference-in-means estimator.

2. In this dataset, what is the average number of new (or repaired) drinking water facilities in villages with a female politician? Please answer with a full sentence. (10 points)

R code:

```
## compute mean of water in villages with female politicians
mean(india$water[india$female==1])
## [1] 23.99074
```

(Recall: we use [] to subset a variable; inside the square brackets, we specify the criterion of selection. For example, we can use the relational operator == to set a logical test; only the observations for which the logical test are true will be extracted. In the case above, we extract from *water* the observations that belong to the villages that were assigned to have a female politician (for which *female* equals 1) and then calculate the mean of those observations.)

Answer: The average number of new (or repaired) drinking water facilities in villages with a female politician is 24 facilities.

3. What is the average number of new (or repaired) drinking water facilities in villages with a male politician? Please answer with a full sentence. (10 points)

R code:

```
## compute mean of water in villages with male politicians
mean(india$water[india$female==0])
## [1] 14.73832
```

Answer: The average number of new (or repaired) drinking water facilities in villages with a male politician is 15 facilities.

4. What is the estimated average causal effect of having a female politician on the number of new (or repaired) drinking water facilities? Please provide a full substantive answer (make sure to include the assumption, why the assumption is reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (25 points)

R code:

```
## compute the difference in means estimator
mean(india$water[india$female==1]) - #average outcome for treatment group MINUS
  mean(india$water[india$female==0]) #average outcome for control group
## [1] 9.252423
```

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

- *What's the assumption?* We assume that the villages that were assigned to have a female politician (the treatment group) are comparable to the villages that were assigned to NOT have a female politician (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 is the assumption reasonable?* Because the female politicians were assigned at random OR because the data come from a randomized experiment. (Recall: Random treatment

assignment makes the treatment and control groups on average identical to each other in all observed and unobserved pre-treatment characteristics.)

- *What's the treatment?* Having a female politician.

- *What's the outcome?* Number of new (or repaired) drinking water facilities.

- *What's the direction, size, and unit of measurement of the average causal effect?* An increase of 9 facilities, 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 difference-in-means estimator = average number of facilities in villages with female politician - average number of facilities in villages with male politician = 24 facilities - 15 facilities = 9 facilities.)

Full answer: Assuming that the villages that were randomly assigned to have a female politician were comparable to the villages that were randomly assigned to NOT have a female politician (a reasonable assumption since the female politicians were assigned at random), we estimate that having a female politician increases the number of new or repaired drinking water facilities by 9 facilities, on average.

(Note: the following statement is not the correct answer because it doesn't make a causal claim; it only makes an observation: Assuming that the villages that were randomly assigned to have a female politician were comparable to the villages that were randomly assigned to NOT have a female politician (a reasonable assumption since the female politicians were assigned at random), we estimate that villages with a female politician have 9 more new or repaired drinking water facilities, on average, than villages with a male politician.)