Do Women Promote Different Policies than Men? Part II: Computing and Interpreting Means (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 since random assignment

Table 1: Variables in "india.csv"

In this problem set, we practice how to compute and interpret means, among other things.

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. Use the function mean() to calculate the average of the variable *female*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)

   R code:

   ```
   mean(india$female) # calculates the average of female
   ## [1] 0.3354037
   ```

   (Recall: We use $ to access a variable within a dataframe. To its left, we specify the name of the object where the dataframe is stored, india in this case; to its right, we specify the name of the variable, female in this case. Also, we do not use quotes around the names of functions, the names of objects, or the names of elements within an object such as variables.)

Answer: 34% of the villages in the experiment were randomly assigned to have a female politician. (Note: The word "average" does not appear in the interpretation because the mean of a binary variable should be interpreted as the proportion of the observations that have the characteristic identified by the variable. Here, *female* is binary and identifies the villages that were assigned to have a female politician. The unit of measurement is %, after multiplying the output by 100. Here: 0.34×100=34%.)

2. Use the function mean() to calculate the average of the variable *water*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)

   R code:

```
mean(india$water) # calculates the average of water
## [1] 17.84161
```

   Answer: After the randomization of politicians, the average number of new (or repaired) drinking water facilities per village is 18 facilities (Note: the word "average" appears in the interpretation of this mean. Recall: The mean of a non-binary variable should be interpreted as an average and in the same unit of measurement as the variable itself, which is facilities in this case.)

3. If we wanted to estimate the average causal effect of having a female politician on the number of new (and repaired) drinking water facilities: (10 points)
   a. What would be the treatment variable? Please just provide the name of the variable
   b. What would be the outcome variable? Please just provide the name of the variable

      Answer: (a) *female*, (b) *water*.

4. If we wanted to estimate the average causal effect of having a female politician on the number of new (and repaired) irrigation facilities: (10 points)
   a. What would be the treatment variable? Please just provide the name of the variable
   b. What would be the outcome variable? Please just provide the name of the variable

      Answer: (a) *female*, (b) *irrigation*.

5. In both analyses above: (10 points)
   a. What would be the treatment group?
   b. What would be the control group?

      Answer: (a) the villages that were randomly assigned to have a female politician, (b) the villages that were randomly assigned to NOT have a female politician. (Note: the unit of observation is villages so both of these groups are composed of villages; the treatment group are the villages that were assigned to receive the treatment—having a female politician; the control group are the villages that were assigned to NOT receive the treatment.)