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.

Do Women Promote Different Policies than Men? Part VI: Effect on Irrigation Facilities

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 how to (1) estimate an average treatment effect using data from a randomized experiment, (2) determine whether the estimated average treatment effect is statistically significant at the 5% level, and (3) determine the internal and external validity of the study.

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. What is the estimated average casual effect of having a female politician on the number of new (or repaired) irrigation facilities?

    a. Fit a linear model to the data in such a way that the estimated slope coefficient is equivalent to the difference-in-means estimator you are interested in and store the fitted model in an object called *fit* (R code only). (2.5 points)

    b. What is the estimated slope coefficient, $\widehat{\beta}$? (2.5 points)

    c. Now, let's answer the question: What is the estimated average treatment effect? 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) (10 points)

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.

2. Is the effect statistically significant at the 5% level?
    a. Let's start by specifying the null and alternative hypotheses. Please provide both the mathematical notations and their meaning. (5 points)
    b. What is the value of the observed test statistic, z^{obs}? (Hint: the code summary()$coeff might be helpful here.) (2.5 points)
    c. What is the associated p-value? (2.5 points)
    d. Now, let's answer the question: Is the effect statistically significant at the 5% level? Please provide your reasoning. (5 points)
3. Can we interpret the estimated effect as causal? In other words, how strong is the internal validity of this study? Have the researchers accurately measured the average causal effect on the sample of villages that were part of the study? Please explain your reasoning. (10 points)
4. Can we generalize the results? In other words, how strong is the external validity of this study? Please explain your reasoning and be specific about what population you think the findings can or cannot be generalized to. (10 points)