Problem set 3: RCTs, matching, and inverse probability weighting

W. Hunter Giles

September 10th, 2022

Table of Contents

# Program overview

The metropolitan Atlanta area is interested in helping residents become more environmentally conscious, reduce their water consumption, and save money on their monthly water bills. To do this, Fulton, DeKalb, Gwinnett, Cobb, and Clayton counties have jointly initiated a new program that provides free rain barrels to families who request them. These barrels collect rain water, and the reclaimed water can be used for non-potable purposes (like watering lawns and gardens). Officials hope that families that use the barrels will rely more on rain water and will subsequently use fewer county water resources, thus saving both the families and the counties money.

Being evaluation-minded, the counties hired an evaluator (you!) before rolling out their program. You convinced them to fund and run a randomized controlled trial (RCT) during 2018, and the counties rolled out the program city-wide in 2019. You have two datasets: barrels\_rct.csv with data from the RCT, and barrels\_obs.csv with observational data from self-selected participants.

These datasets contain the following variables:

* id: A unique ID number for each household
* water\_bill: The family’s average monthly water bill, in dollars
* barrel: An indicator variable showing if the family participated in the program
* barrel: A 0/1 numeric version of barrel
* yard\_size: The size of the family’s yard, in square feet
* home\_garden: An indicator variable showing if the family has a home garden
* home\_garden\_num: A 0/1 numeric version of home\_garden
* attitude\_env: The family’s self-reported attitude toward the environment, on a scale of 1-10 (10 meaning highest regard for the environment)
* temperature: The average outside temperature (these get wildly unrealistic for the Atlanta area; just go with it)

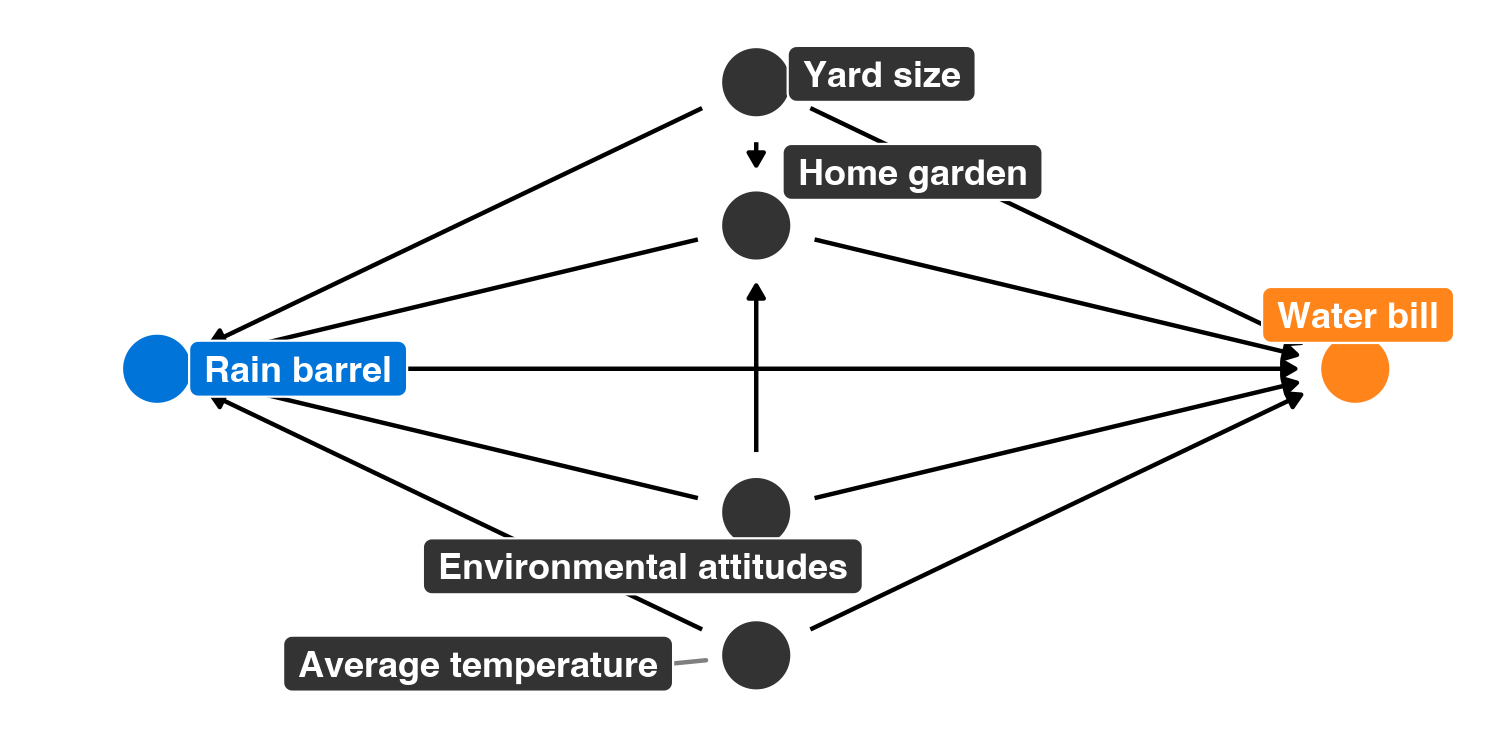
# Your goal

Your task in this problem set is to analyze these two datasets to find the causal effect (or average treatment effect (ATE)) of this hypothetical program.

***Follow these two examples from class as guides:***

* [RCTs](https://evalsp22.classes.andrewheiss.com/example/rcts/)
* [Matching and IPW](https://evalsp22.classes.andrewheiss.com/example/matching-ipw/)

As a reference, Figure 1 shows the DAG for the program:



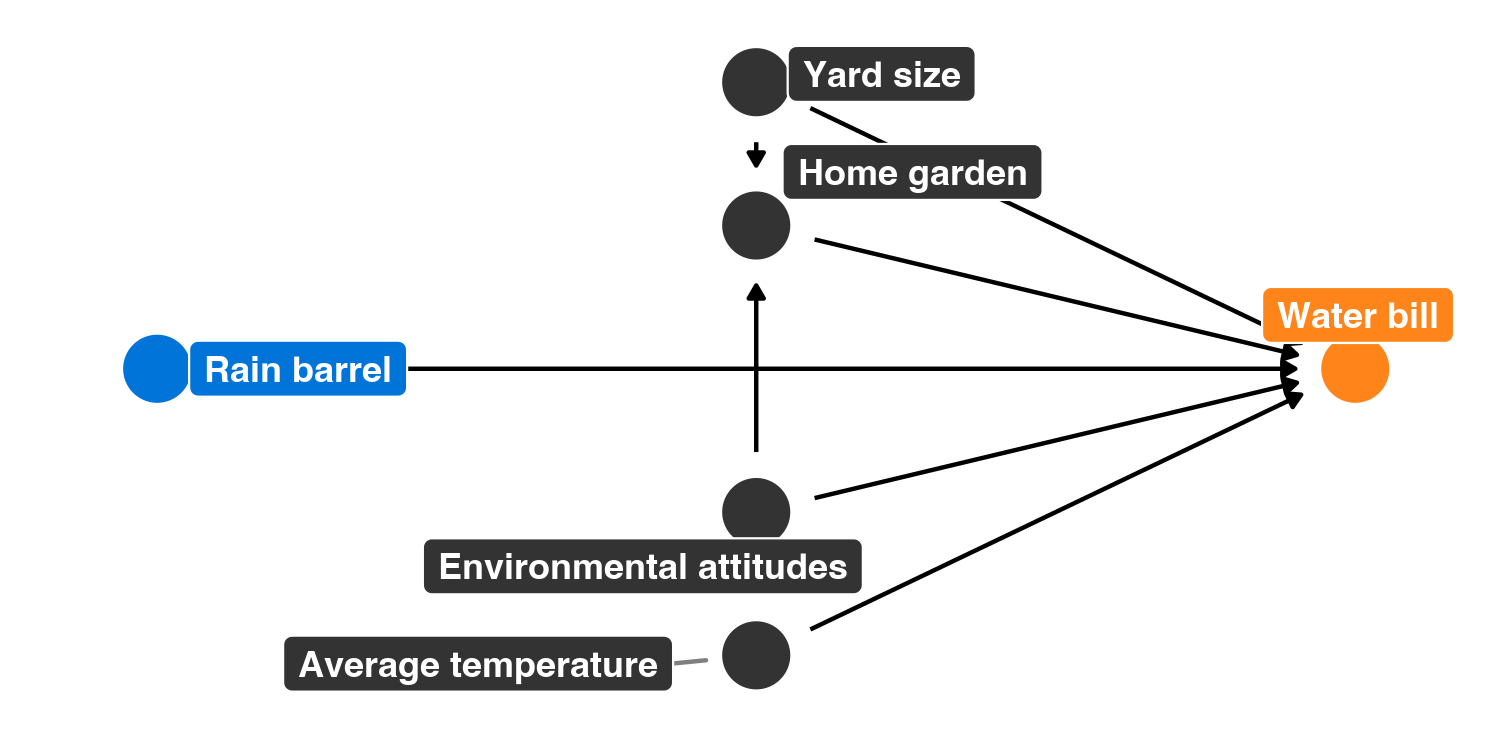
Rain barrel program DAG

library(tidyverse)  
library(broom)  
library(patchwork)  
library(MatchIt)  
library(modelsummary)  
  
  
barrels\_rct <- read\_csv("data/barrels\_rct.csv") %>%   
 # This makes it so "No barrel" is the reference category  
 mutate(barrel = fct\_relevel(barrel, "No barrel"))  
  
barrels\_obs <- read\_csv("data/barrels\_observational.csv") %>%   
 # This makes it so "No barrel" is the reference category  
 mutate(barrel = fct\_relevel(barrel, "No barrel"))

# 1. Finding causation from a randomized controlled trial

## Modified DAG

You remember from PMAP 8521 that when running an RCT, you can draw the DAG for the program like this (Figure 2). **Why?**



Rain barrel program DAG as an RCT

## Check balance

**Discuss the sample size for the RCT data and how many people were assigned to treatment/control. Are you happy with this randomization?**

# Check for balance of numbers in the treatment and control groups  
barrels\_rct %>% group\_by(barrel) %>%   
 summarise(cnt = n()) %>% mutate(freq = (cnt / sum(cnt)))

## # A tibble: 2 × 3  
## barrel cnt freq  
## <fct> <int> <dbl>  
## 1 No barrel 221 0.448  
## 2 Barrel 272 0.552

The treatment and control group have roughly the same number of observations in them. This balance will help the authenticity of the RCT. However, this is a city wide policy, so I would like more observations than 493.

**Check the balance of the main pre-treatment characteristics. Are you happy with the balance?**

Home garden frequency is about the some in treatment and control.

# Insert as many chunks as you need to check the balance of the RCT across  
# different pre-treatment characteristics like yard size, home garden,  
# environmental attitudes, and average temperature.  
barrels\_rct %>% group\_by(barrel,home\_garden) %>% summarise(cnt = n()) %>% mutate(freq = cnt / sum(cnt))

## # A tibble: 4 × 4  
## # Groups: barrel [2]  
## barrel home\_garden cnt freq  
## <fct> <chr> <int> <dbl>  
## 1 No barrel Home garden 59 0.267  
## 2 No barrel No home garden 162 0.733  
## 3 Barrel Home garden 56 0.206  
## 4 Barrel No home garden 216 0.794

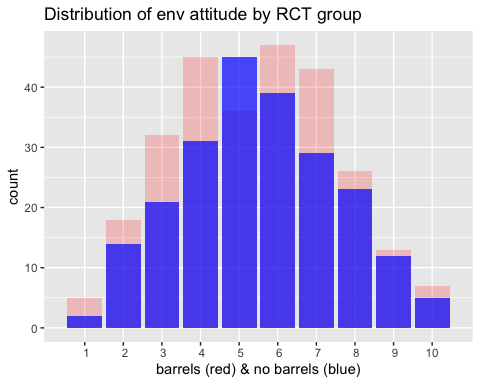
environmental attitide is relatively balanced

barrels\_rct %>% group\_by(barrel) %>% summarise(  
 mean= mean(attitude\_env))

## # A tibble: 2 × 2  
## barrel mean  
## <fct> <dbl>  
## 1 No barrel 5.52  
## 2 Barrel 5.42

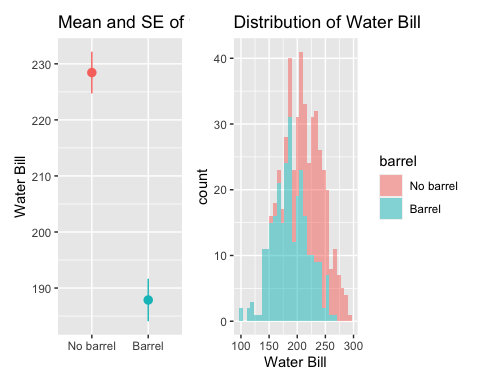
This seems balanced to me.

ggplot(barrels\_rct) +  
 geom\_bar(data = filter(barrels\_rct, barrel == "Barrel"), mapping = aes(x=attitude\_env), fill = "red", alpha = 0.2) +   
 geom\_bar(data = filter(barrels\_rct, barrel == "No barrel"), mapping = aes(x=attitude\_env), fill = "blue", alpha = .7) +  
 scale\_x\_discrete(limits = seq(1,10)) +   
 labs(x = "barrels (red) & no barrels (blue)", title = "Distribution of env attitude by RCT group")



It looks like the distributions of environmental attitude are about the same between the treatment and control group, however, there seems to be more people in the barrels category.

waterbill\_summary <- ggplot(barrels\_rct, mapping = aes(x = barrel, y = water\_bill, color = barrel)) +  
 stat\_summary(geom = "pointrange", fun.data = "mean\_se", fun.args = list(mult = 1.96)) +  
 guides(color = "none") +  
 labs(title = "Mean and SE of water bill", x = NULL, y = "Water Bill")  
  
waterbill\_dis <- ggplot(barrels\_rct, aes(x=water\_bill, fill=barrel)) +  
 geom\_histogram(alpha = .5) +   
 labs(title = "Distribution of Water Bill", x = "Water Bill")  
  
waterbill\_summary + waterbill\_dis



There is a difference in means which is ok because water\_bill is the dependent variable. The distributions are both relatively normal.

## Estimate difference

**What is the effect of the program on participants’ water bills? How credible is this result?**

# Insert as many chunks as you need to measure the difference in water bills for  
# those in the program and those not in the program.  
df <- barrels\_rct %>% group\_by(barrel) %>% summarise(mean = mean(water\_bill))  
abs(df$mean[1] - df$mean[2])

## [1] 40.57346

# 2. Finding causation from observational data

## Naive difference in means

**As a baseline measure, calculate the average difference in water bills for those in the program and those not in the program using the observational data. How much of an effect is there? How credible is this result? Why?**

# Find the naive difference in means  
  
naive\_model <- lm(water\_bill ~ barrel\_num,   
 data = barrels\_obs)  
tidy(naive\_model)

## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 225. 1.07 211. 0   
## 2 barrel\_num -29.9 1.67 -17.8 1.61e-63

## Adjustment with Mahalanobis nearest-neighbor matching

**Use matching to make the proper adjustments and close the backdoor effects from the four main confounders: yard\_size, home\_garden, attitude\_env, and temperature**

# Insert as many chunks as you need to close the backdoors by using nearest  
# neighbor matching with Mahalanobis distance.  
  
matched\_data <- matchit( barrel\_num ~ yard\_size + home\_garden\_num + attitude\_env + temperature,   
 data = barrels\_obs,  
 method = "nearest",  
 distance = "mahalanobis",  
 replace = TRUE)  
  
matched\_data\_for\_real <- match.data(matched\_data)  
  
matched\_model <- lm(water\_bill ~ barrel\_num,   
 data = matched\_data\_for\_real,  
 weights = weights)  
tidy(matched\_model)

## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 234. 1.68 140. 0   
## 2 barrel\_num -39.4 2.12 -18.6 1.46e-64

## Adjustment with inverse probability weighting

**Use inverse probability weighting to make the proper adjustments and close the backdoor effects from the four main confounders: yard\_size, home\_garden, attitude\_env, and temperature**

# Insert as many chunks as you need to close the backdoors by using propensity  
# scores and inverse probability weighting.  
logit\_model <- glm(barrel\_num ~ yard\_size + home\_garden\_num + attitude\_env + temperature,  
 data = barrels\_obs,  
 family = binomial(link = "logit"))  
  
barrel\_ipw <- augment\_columns(logit\_model,   
 barrels\_obs,   
 type.predict = "response") %>%   
 mutate(weight = (barrel\_num / .fitted) + ((1-barrel\_num)/(1-.fitted)))  
  
  
ipw\_model <- lm(water\_bill ~ barrel\_num,   
 data = barrel\_ipw,  
 weights = weight)  
  
tidy(ipw\_model)

## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 228. 1.19 192. 0   
## 2 barrel\_num -39.1 1.67 -23.4 1.68e-100

# 3. Comparing results

You just calculated a bunch of ATEs using experimental and observational data. **Put them all in a single table here:**

# Table with all model results side-by-side  
modelsummary(list(  
 "Naive" = naive\_model,  
 "Matched w/ weights" = matched\_model,  
 "IPW truncated" = ipw\_model))

|  | Naive | Matched w/ weights | IPW truncated |
| --- | --- | --- | --- |
| (Intercept) | 224.800 | 234.371 | 228.214 |
|  | (1.068) | (1.678) | (1.189) |
| barrel\_num | -29.860 | -39.431 | -39.050 |
|  | (1.674) | (2.118) | (1.670) |
| Num.Obs. | 1241 | 805 | 1241 |
| R2 | 0.204 | 0.301 | 0.306 |
| R2 Adj. | 0.204 | 0.301 | 0.306 |
| AIC | 11881.0 | 7758.9 | 12016.2 |
| BIC | 11896.3 | 7772.9 | 12031.6 |
| Log.Lik. | -5937.482 | -3876.430 | -6005.090 |
| F | 318.160 | 346.456 | 547.025 |
| RMSE | 28.97 | 29.06 | 41.69 |

**Which estimates do you believe? Why? Would observational ATEs be sufficient to prove program effect in this case? Why or why not? Should this program be rolled out throughout Georgia? Why or why not?**

From the 4 estimates above, the most sophisticated model is the RCT. Since the treatment and control group are randomly selected, I can assume that the treatment effect is unbiased. This is because all other confounding variables should cancel out. The IPW and matching only control for yard size, home garden, environmental attitude, and temperature. However, it is still a good robustness check. From the data, we can see that the barrel program significantly decreases water bills. Since a RCT model is used, we have a good case for optimal external validity, so this program should be used.