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

W. Hunter Giles

September 10th, 2022

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.

- RCTs
- Matching and IPW

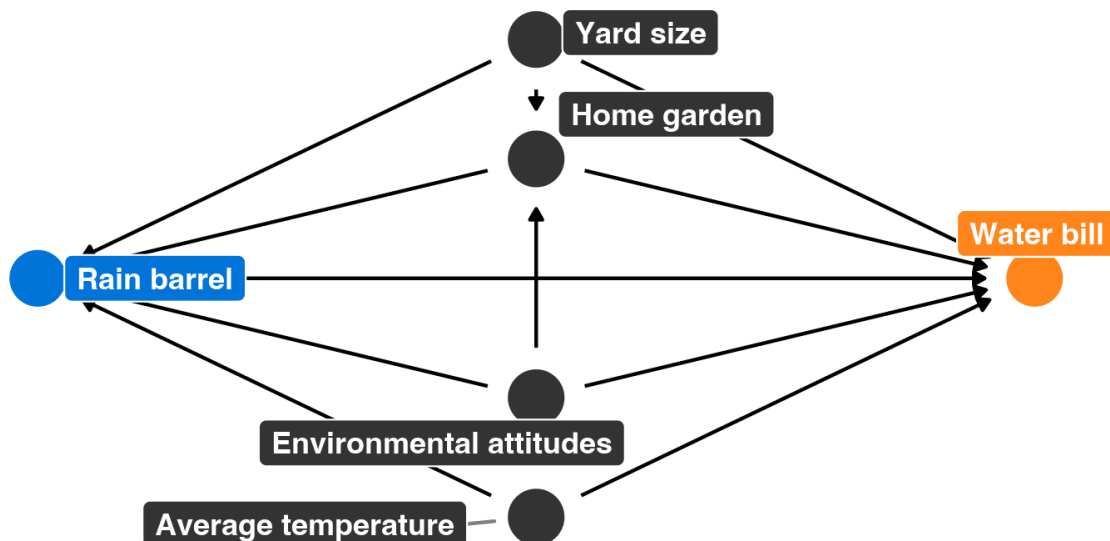As a reference, Figure 1 shows the DAG for the program:

Figure 1: Rain barrel program DAG

```
library(tidyverse)
library(broom)
library(patchwork)
library(MatchIt)
```

```r
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?**

The treatment is randomly selected, so there is nothing biasing the selection. This is way there are no arrows pointed at the *Rain barrel* variable.
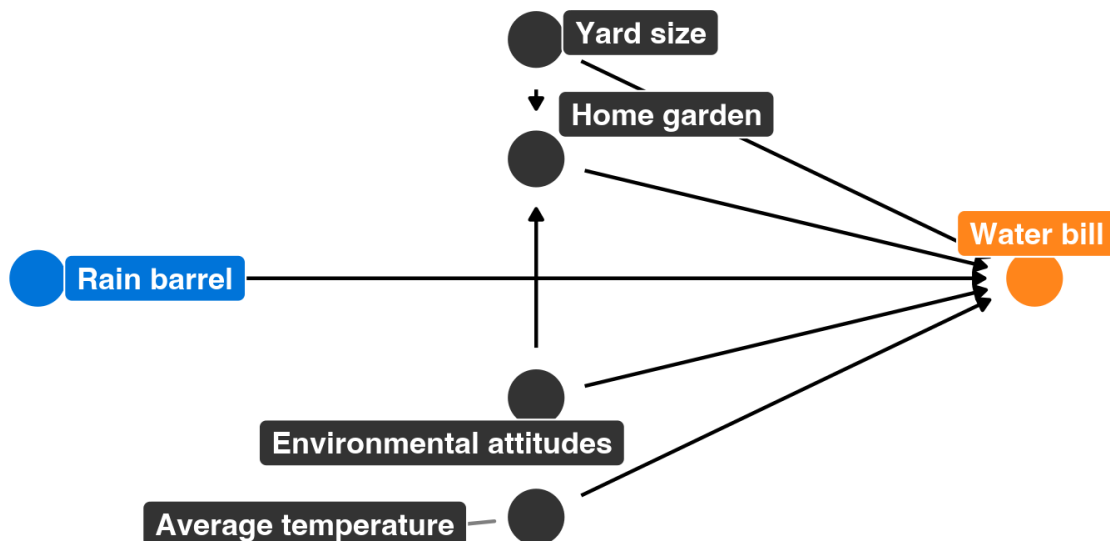

Figure 2: Rain barrel program DAG as an RCT

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?

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 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 x 3
##   barrel       cnt  freq
##   <fct>      <int> <dbl>
## 1 No barrel    221 0.448
## 2 Barrel       272 0.552
```

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 x 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 attitude is relatively balanced.

```
barrels_rct %>% group_by(barrel) %>% summarise(cnt = n(),
                                     mean= mean(attitude_env)) %>% mutate(freq = cnt /sum(cnt)
```

```
## # A tibble: 2 x 4
##   barrel       cnt  mean  freq
##   <fct>      <int> <dbl> <dbl>
## 1 No barrel    221  5.52 0.448
## 2 Barrel       272  5.42 0.552
```

This seems balanced to me.

```
ggplot(barrels_rct) +
  geom_bar(data = filter(barrels_rct, barrel == "Barrel"), mapping = aes(x=attitude_env), fill = "red",
  geom_bar(data = filter(barrels_rct, barrel == "No barrel"), mapping = aes(x=attitude_env), fill = "blu
  scale_x_discrete(limits = seq(1,10)) +
  labs(x = "barrels (red) & no barrels (blue)", title = "Distribution of env attitude by RCT group")
```

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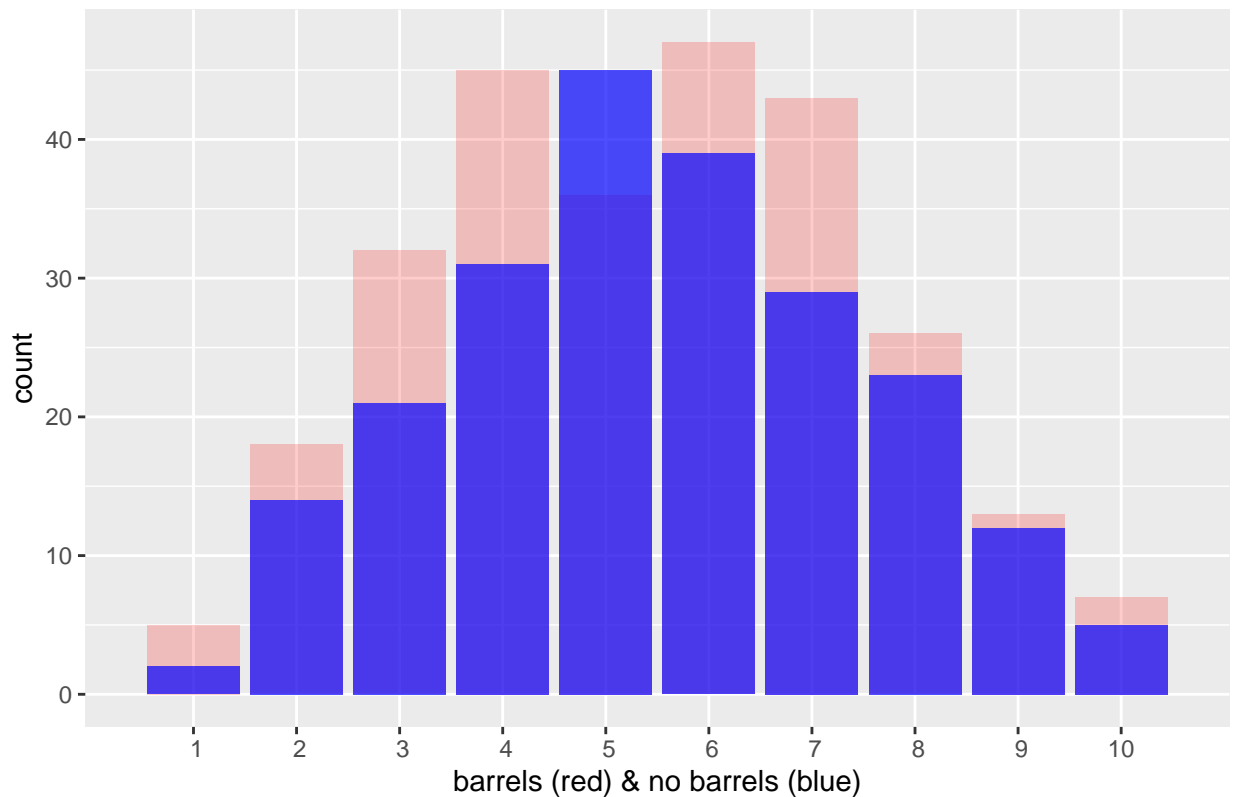
It looks like the distributions of environmental attitude are about the same between the treatment and control group, however, there seems to be more people in the barrels category.

```r
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
```
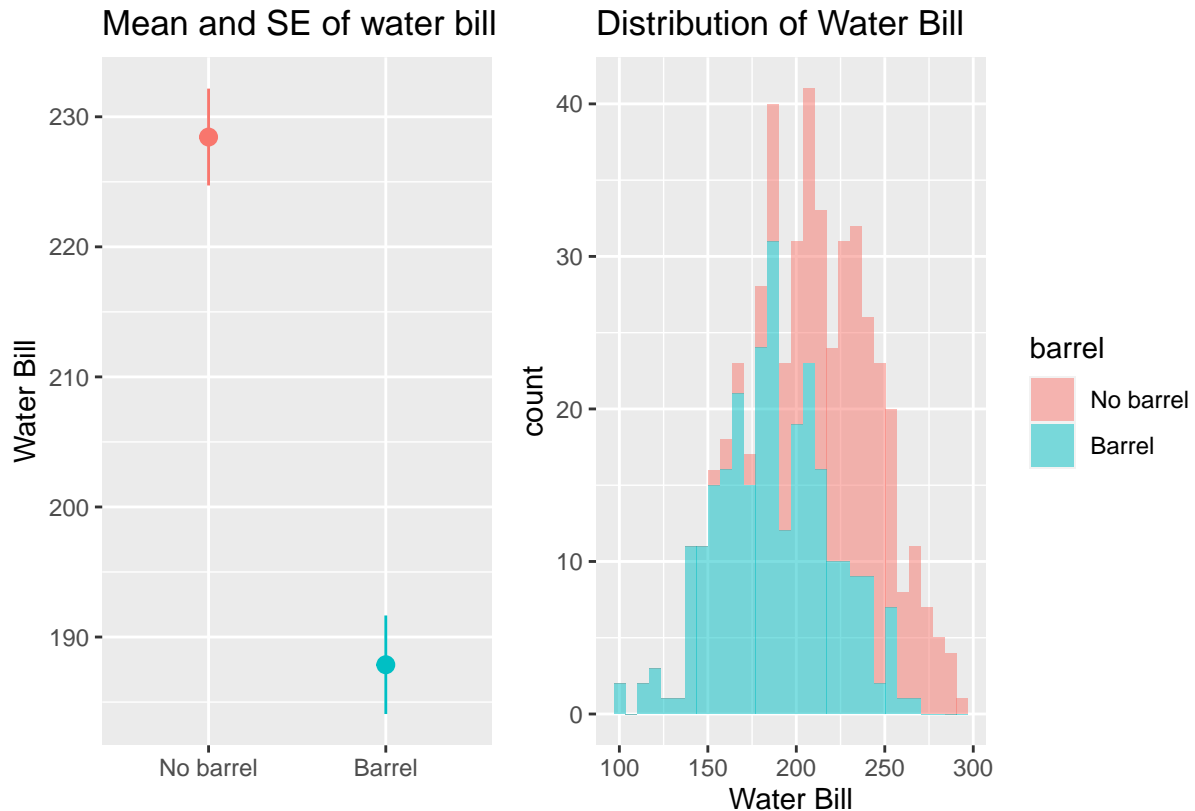
There is a difference in means which is ok because water_bill is the dependent variable. The distributions are both relatively normal.

Estimate difference

What is the effect of the program on participants' water bills? How credible is this result?
The RCT results show that the treatment causes a 40 dollar decrease in families' water bills, on average. These results are credible because the treatment effect is unbiased.

```
# 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?

The naive shows a $29 decrease in water bill cost. However, this is biases because confounding variables are not being controlled for. This means that there is likely something influencing a persons selection into the treatment.

```
# Find the naive difference in means

naive_model <- lm(water_bill ~ barrel_num,
                  data = barrels_obs)
tidy(naive_model)
```

```
## # A tibble: 2 x 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 x 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,
```

	Naive	Matched w/ weights	IPW truncated
(Intercept)	224.800	234.371	228.214
	(1.068)	(1.678)	(1.180)
barrel_num	−29.860	−39.431	−38.454
	(1.674)	(2.118)	(1.664)
Num.Obs.	1241	805	1241
R2	0.204	0.301	0.301
R2 Adj.	0.204	0.301	0.301
AIC	11 881.0	7758.9	12 000.3
BIC	11 896.3	7772.9	12 015.7
Log.Lik.	−5937.482	−3876.430	−5997.145
F	318.160	346.456	533.896
RMSE	28.97	29.06	41.39

```
                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)))

# removing weight score outliers
barrel_ipw <- mutate(barrel_ipw, weight = ifelse(weight > 8,8,weight))

ipw_model <- lm(water_bill ~ barrel_num,
                data = barrel_ipw,
                weights = weight)

tidy(ipw_model)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    228.       1.18     193.  0
## 2 barrel_num     -38.5      1.66     -23.1 1.64e-98
```

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))
```

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.

One interesting note is in the ipw model when adjusting for weight outliers, the treatment effect (38.45) is farther away from the RCT treatment effect (40.57). When I dont adjust for the weight outliers, the ipw results (39.0) are closer to the RCT.