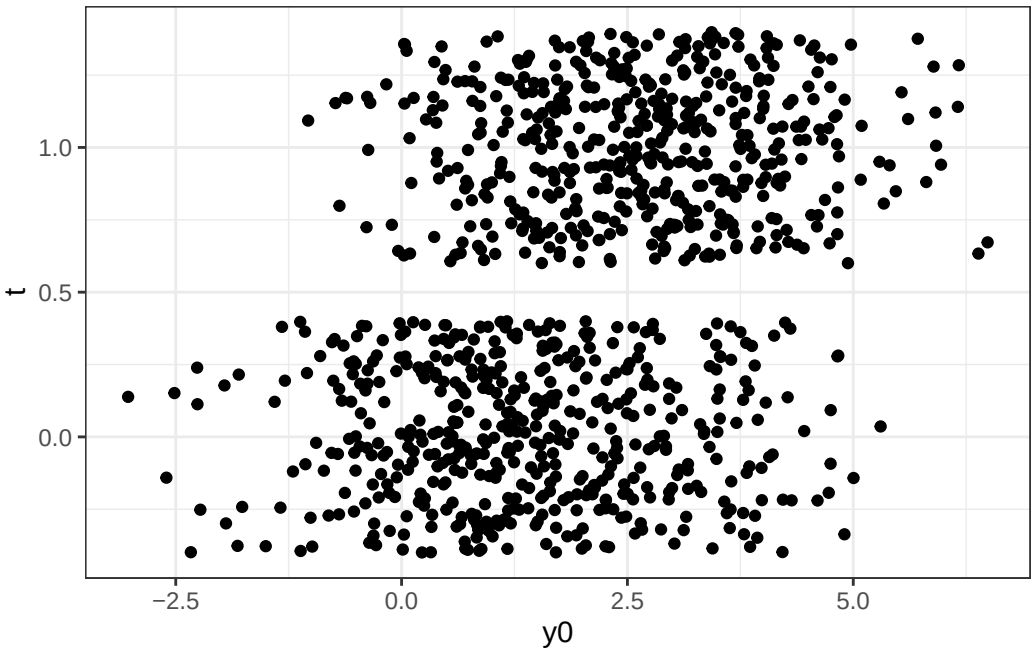
Quant2 Lab1 Full

2026-01-29

Potential Outcomes Data Simulation

```
pacman::p_load(tidyverse, here, knitr, kableExtra)

set.seed(123)
N <- 1000

# 2 random covariates
x1 <- runif(N, 0, 1)
x2 <- rnorm(N, 0, 0.5)

# Some noise
e <- rnorm(N, 0, 1)

# Treatment effect
d <- rnorm(N, 1, 1)

# Potential outcomes
y0 <- x1 * 4 + x2 + e
y1 <- y0 + d

# Treatment assignment, confounded by x1
ts <- rbinom(n = N, size = 1, prob = x1)

df <- data.frame(
  x1=x1,
  x2=x2,
  x3=rnorm(N, 1, 2),
  y0=y0,
```

```
  y1=y1,
  t=ts,
  y=ts * y1 + (1-ts) * y0 #observed
)
df %>% ggplot(aes(x=y0, y=t)) + geom_point(position='jitter') + theme_bw()
```




```
# df %>% write_tsv(here('lab01','thescience.tsv'))
```

- How does the observed outcome `y` relate to the treatment `t` and the potential outcomes `y0` and `y1`?

```
df %>% select(y0, y1, t, y) %>% tail
```

```
            y0        y1 t         y
995  0.6156253 0.5007992 0 0.6156253
996  4.1522994 4.9236784 1 4.9236784
997  4.4556111 5.3740212 1 5.3740212
998  1.8652038 2.1326373 1 2.1326373
999  4.8341564 4.4515235 0 4.8341564
1000 0.9952832 3.8340798 0 0.9952832
```

- Calculate the difference in means between the treated and the untreated.

```r
mean(filter(df, t == 1)$y) - mean(filter(df, t == 0)$y)
```

```
[1] 2.257254
```

- Calculate the true global average treatment effect

```r
mean(df$y1 - df$y0)
```

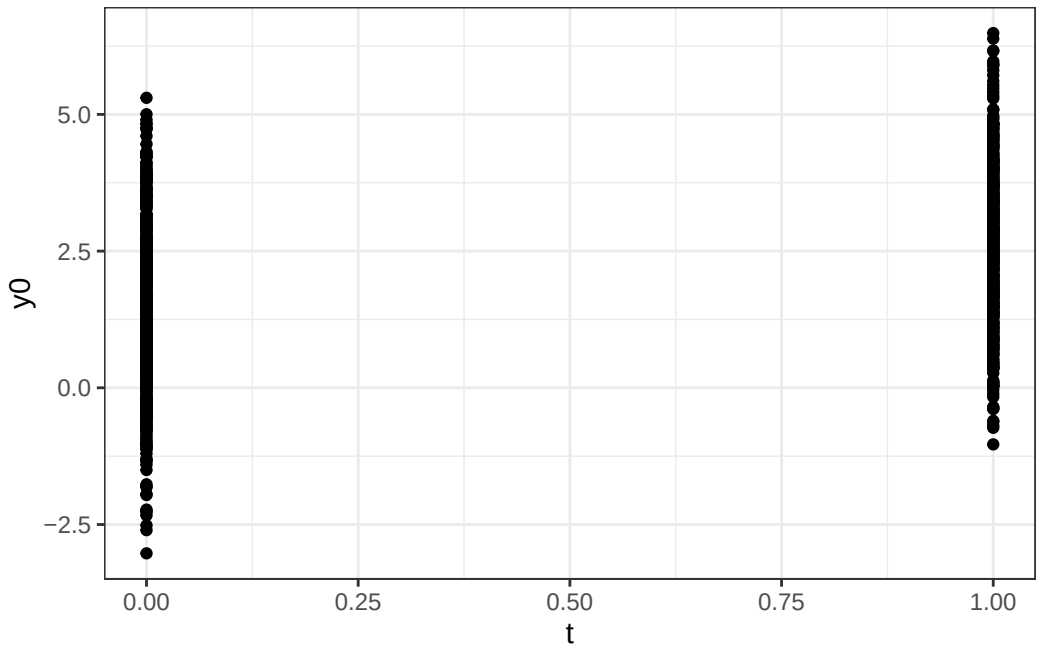
```
[1] 1.000603
```

- Why they are different?

```r
# Selection bias because there is a difference in the potential outcomes of
    the treatment vs. control
mean(filter(df, t == 1)$y0) - mean(filter(df, t == 0)$y0)
```

```
[1] 1.256952
```

```r
# We can see this graphically by plotting the data
df %>% ggplot(aes(x =t , y = y0)) + geom_point() + theme_bw()
```



```r
# Or we can observe this by noting that treatment and potential outcomes are
    correlated
cor(df$t, df$y0)
```

```
[1] 0.3990395
```

```r
# Selection bias wrt ATE?
rho_i <- df$y1 - df$y0
mean(rho_i[df$t == 1]) - mean(rho_i[df$t == 0])
```

```
[1] -0.0006006687
```

Exercises

- What is the ATE vs the ATC and ATT? How would we calculate these from the science?

```r
# ATE (Average treatment effect - for everyone)
mean(df$y1 - df$y0)
```

```
[1] 1.000603
```

```r
## Same thing with fancy code
with(df, mean(y1) - mean(y0))
```

```
[1] 1.000603
```

```r
# ATC (Average treatment on control)
mean(filter(df, t==0)$y1 - filter(df, t==0)$y0)
```

```
[1] 1.000903
```

```r
# with(filter(df, t==0), mean(y1) - mean(y0))

# ATT (Average treatment on treated)
mean(filter(df, t==1)$y1 - filter(df, t==1)$y0)
```

```
[1] 1.000303
```

```
# with(filter(df, t==1), mean(y1) - mean(y0))
```

- Which of x1, x2, and x3 are associated with treatment assignment? Are they potentially confounders for identifying the ATE?

```
kable(
  tibble(
  variable = c("$x_1$", "$x_2$", "$x_3$"),
  diff_in_means_T = c(
    mean(df$x1[df$t == 1]) - mean(df$x1[df$t == 0]),
    mean(df$x2[df$t == 1]) - mean(df$x2[df$t == 0]),
    mean(df$x3[df$t == 1]) - mean(df$x3[df$t == 0])
  ),
  cor_with_y0 = c(
    cor(df$x1, df$y0),
    cor(df$x2, df$y0),
    cor(df$x3, df$y0)
  )
  ),
  digits = 3,
  col.names = c(
    "Variable",
    "Diff. in Means by T",
    "Correlation with $Y_0$"
  )
)
```

Variable	Diff. in Means by T	Correlation with Y_0
x_1	0.335	0.697
x_2	0.013	0.344
x_3	0.248	0.039

Fixing the ATE estimation

- You get to play omnipotent being! Create an alternate universe (ie, a new treatment assignment and new outcome variable) such that the difference in means between the treated and the untreated can be reliably estimated.

- Estimate the difference in means and compare it to the true effect.
- Are they different? Why/How?

```
# Assign a random treatment
set.seed(1)
df$treat <- runif(nrow(df), 0, 1) > 0.3
df$outcome <- with(
  df,
  ifelse(treat, y1, y0)
)
# The difference in means is closer to 1 now
mean(filter(df, treat == 1)$outcome) - mean(filter(df, treat == 0)$outcome)
```

```
[1] 1.006773
```

```
mean(df$y1 - df$y0)
```

```
[1] 1.000603
```

```
mean(filter(df, treat==1)$y0) - mean(filter(df, treat==0)$y0)
```

```
[1] 0.01123474
```