# hw2

```
library(tidyverse)
```

## 1

### a

I have some concerns about no interference since no one takes a class in complete isolation and thus students in a given AP Calc class could certainly have an effect on the potential outcomes of the other students in their class. For example, Teddy might befriend Ron who has developed a giant college cheating scheme which would lead to Teddy having a higher college GPA than if he had not take AP Calc with Ron.

I also definitely have concerns about same versions of the treatment as each AP Calc professor, and thus class, is likely different. For example, one student might have Wenny Lainstein, the greatest teacher of this generation (extra credit pls), and another might have Daniel Zhou who shows up to teach 15 minutes late everyday and forces the class to watch soccer everyday instead of learning Calc.

### b

I feel fine about same versions of the treatment in this situation, but I do have concerns about no interference given that all the participants from the study come from the same neighborhood in Manhattan. For example, suppose every member of a family of 5 is chosen for the study. If 4 of the members are in the treatment group and 1 is not, then the family member who was not given the treatment is at far less of a risk of becoming infected, than another member in the community who also was not given the treatment, but spends all of their time around people who are unvaccinated.

**c**

Researchers want to know how effective a special new blue pill is at making people better at causal inference. The participants from the study come from various neighborhoods spread across the United States. The researchers redefine the unit of analysis to be the entire neighborhood, to deal with the fact that no interference is probably not satisfied otherwise.

**2**

**a**

Derive the minimal condition for $\hat{\tau}_{dim}$ to be unbiased for the ATE. What does this condition mean in words?

$$E[\hat{\tau}_{\text{dim}}] - \text{ATE} = \left( E[Y_i(1) \mid D_i = 1] - E[Y_i(1)] \right) - \left( E[Y_i(0) \mid D_i = 0] - E[Y_i(0)] \right)$$

**b**

Derive the minimal condition for $\hat{\tau}_{\text{dim}}$ to be unbiased for the ATT. What does this condition mean in words?

$$E[\hat{\tau}_{\text{dim}}] - \text{ATT} = \left( E[Y_i(1) \mid D_1 = 1] - E[Y_i(0) \mid D_i = 0] \right) - E[Y_i(1) - Y_i(0) \mid D_i = 1]$$

$$= \left( E[Y_i(1) \mid D_1 = 1] - E[Y_i(0) \mid D_i = 0] \right) - \left( E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 1] \right)$$

$$= E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

So we have unbiasedness when

$$E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

In words this means that the expected value of the potential outcome for unit $i$ without treatment is the same regardless of what value of the treatment we condition on.

**c**

Derive the minimal condition for $\hat{\tau}_{\text{dim}}$ to be unbiased for the ATC. What does this condition mean in words?

$$E[\hat{\tau}_{\text{dim}}] - \text{ATC} = \Big( E[Y_i(1) \mid D_1 = 1] - E[Y_i(0) \mid D_i = 0] \Big) - E[Y_i(1) - Y_i(0) \mid D_i = 0]$$

$$= \Big( E[Y_i(1) \mid D_1 = 1] - E[Y_i(0) \mid D_i = 0] \Big) - \Big( E[Y_i(1) \mid D_i = 0] - E[Y_i(0) \mid D_i = 0] \Big)$$

$$= E[Y_i(1) \mid D_i = 1] - E[Y_i(1) \mid D_i = 0]$$

So we have unbiasedness when

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0]$$

In words this means that the expected value for the potential outcome for unit $i$ with treatment is the same regardless of what value of the treatment we condition on.

**d**

In all of the above cases $(Y_i(1), Y_i(0)) \perp\!\!\!\perp$ does indeed imply these conditions, but

**e**

We can see that the minimal condition is met here because

$$E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] = 0 - 0 = 0$$

Of course $(Y_i(1), Y_i(0)) \perp\!\!\!\perp$ does not hold, because both $D$ and $Y_i(0)$ have a dependence on $X$.

**f**

We do in fact see that in this DGP $\hat{\tau}_{dim}$ is unbiased for the ATT

```
dgp_p2 <- function(n) {
  tibble(
    x = runif(n, 0, 1),
```

```r
    d = rbinom(n, 1, 0.25 + 0.5*x),
    y_0 = rnorm(n, 0, 2*x^2 + 1),
    y_1 = rnorm(n, 2, 1)
  )
}

dims <- c()
for(i in 1:1000) {
  data <- dgp_p2(1000)
  dim <- data %>%
    mutate(y = case_when(
      d == 1 ~ y_1,
      d == 0 ~ y_0
    )) %>%
    group_by(d) %>%
    summarise(each = mean(y))

  dims <- c(dims, dim$each[2] - dim$each[1])
}

tibble(
  x = dims
) %>%
  ggplot(aes(x = 1, y = dims)) +
  geom_boxplot()
```
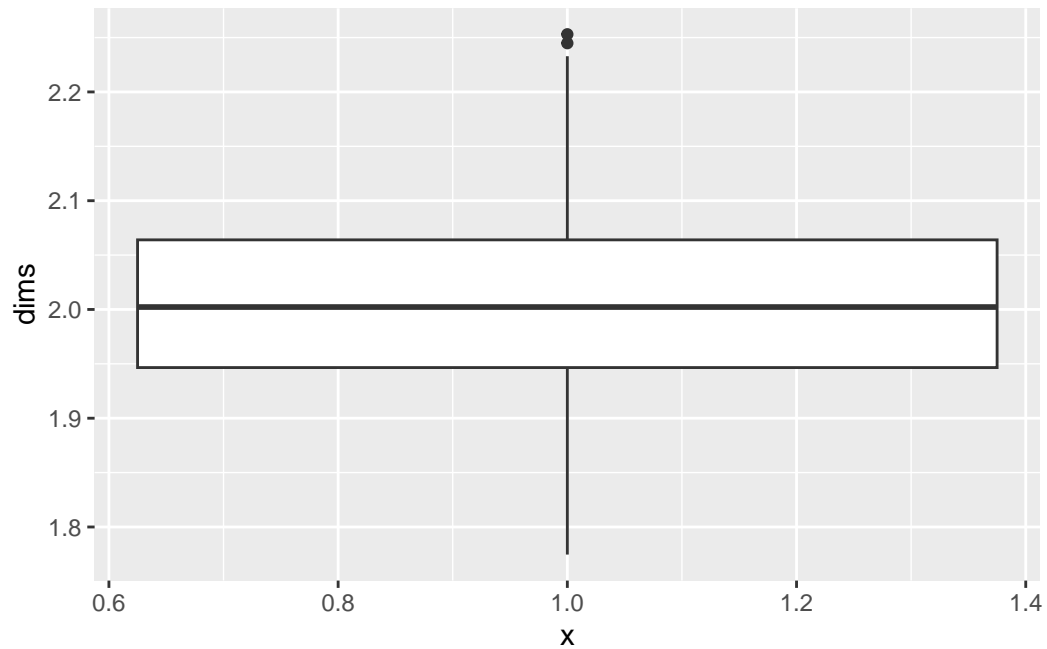
**3**

```r
# helper function to run the simulation in each part
run_sim <- function(n, .f) {
  .f = as_mapper(.f)
  res <- data.frame()

  for(i in 1:1000) {
    data <- .f(1000)
    satc <- data %>%
      filter(d == 0) %>%
      summarise(satc = mean(Y_1 - Y_0)) %>%
      pull()

    satt <- data %>%
      filter(d == 1) %>%
      summarise(satt = mean(Y_1 - Y_0)) %>%
      pull()

    sate <- data %>%
      summarise(sate = mean(Y_1 - Y_0)) %>%
      pull()
```

```
    res <- rbind(res, tibble(satc = satc, satt = satt, sate = sate))
  }

  res <- res %>%
    pivot_longer(cols = 1:3, names_to = "var", values_to = "value")

  return(res)

}
```

**a**

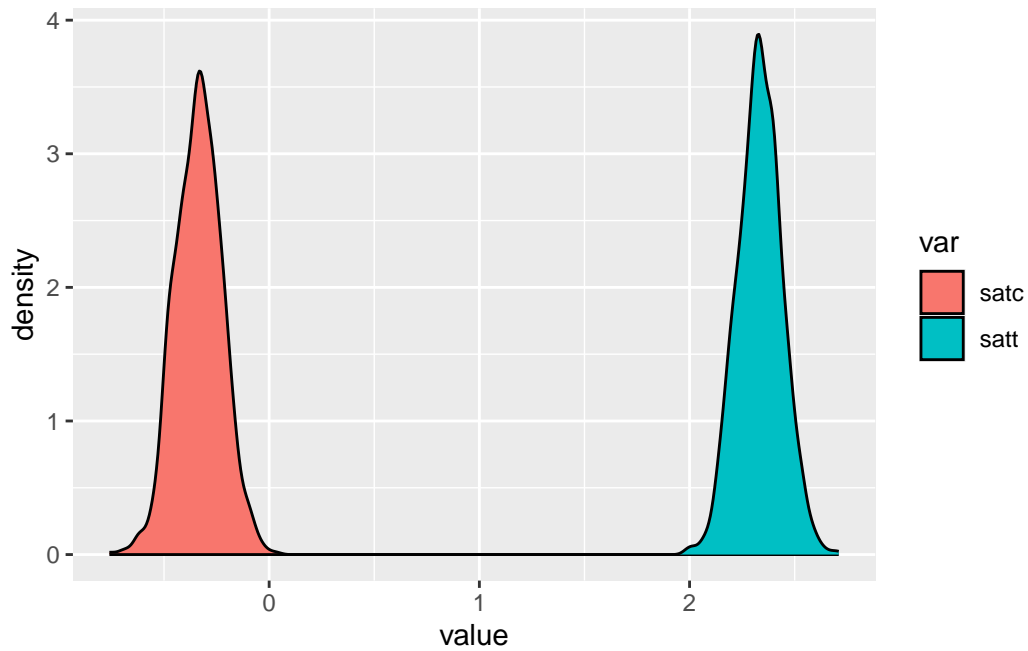I hypothesize that the ATT and ATC will be different

```
dgp_a <- function(n) {
  tibble(
    x = runif(n, -1, 1),
    d = rbinom(n , 1, 0.5 + 0.5*x),
    Y_0 = x + rnorm(n, 0, 1),
    Y_1 = 1 + 5*x + rnorm(n, 0 , 1)
  )
}

a_res <- run_sim(1000, dgp_a)

a_res %>%
  filter(var != "sate") %>%
  ggplot(aes(x = value, fill = var)) +
    geom_density()
```
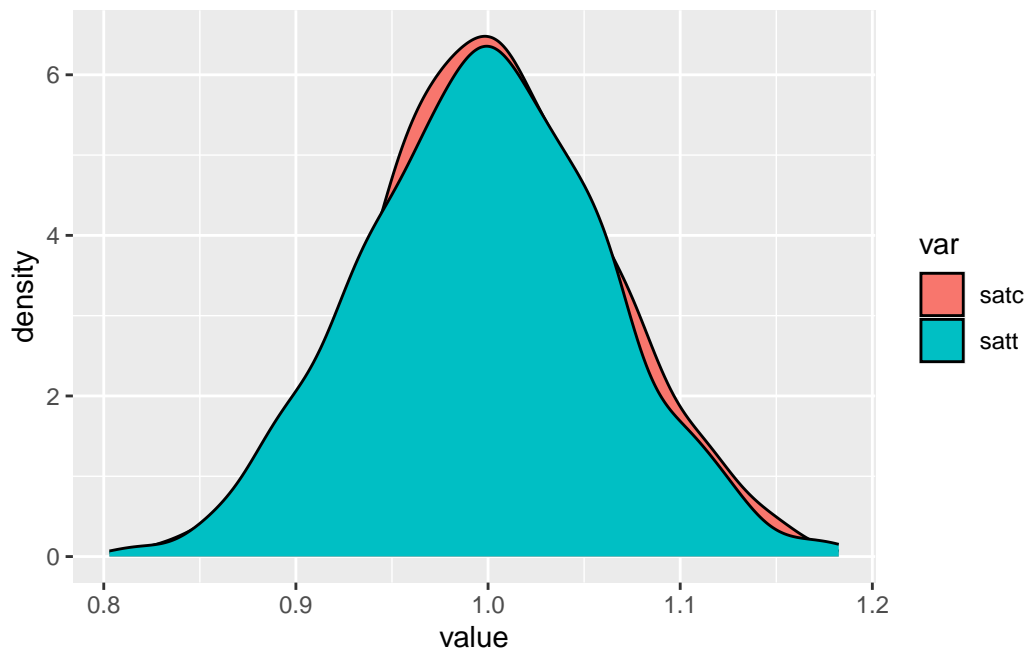
**b**

I hypothesize that the ATT and ATC will be different

```r
dgp_b <- function(n) {
  tibble(
    x = runif(n, -1, 1),
    d = rbinom(n , 1, 0.5 + 0.5*x),
    Y_0 = x + rnorm(n, 0, 1),
    Y_1 = 1 + x + rnorm(n, 0 , 1)
  )
}


b_res <- run_sim(1000, dgp_b)

b_res %>%
  filter(var != "sate") %>%
  ggplot(aes(x = value, fill = var)) +
    geom_density()
```
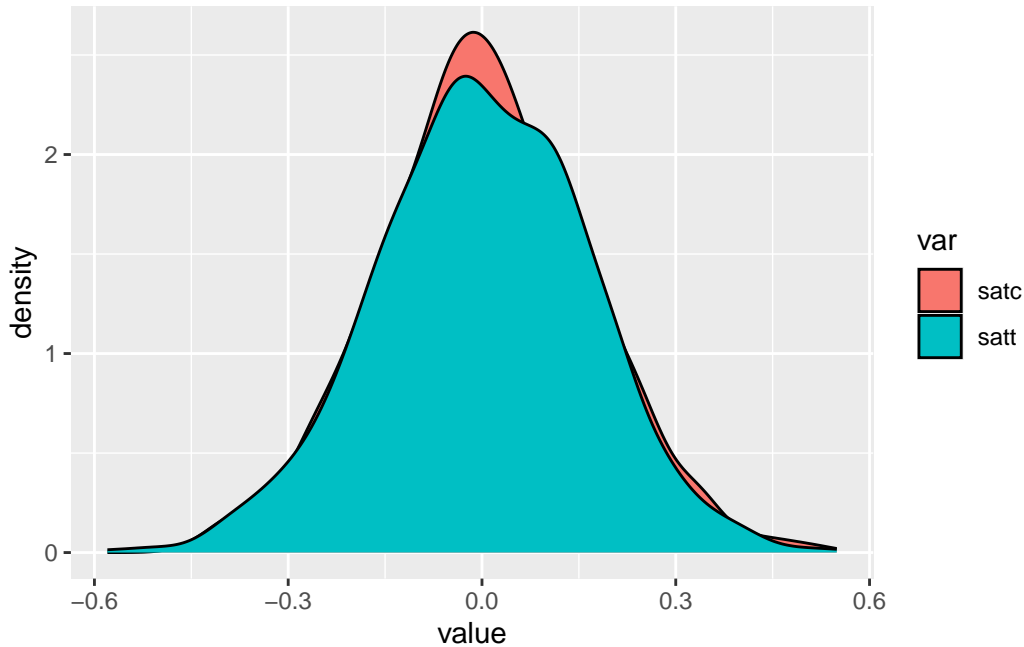
**c**

I hypothesize that the ATT and ATC will be the same

```r
dgp_c <- function(n) {
  tibble(
    x = runif(n, -1, 1),
    d = rbinom(n , 1, 0.5),
    Y_0 = 1 - x + rnorm(n, 0, 1),
    Y_1 = 1 + 5*x + rnorm(n, 0 , 1)
  )
}


c_res <- run_sim(1000, dgp_c)

c_res %>%
  filter(var != "sate") %>%
  ggplot(aes(x = value, fill = var)) +
    geom_density()
```

**d**

The ATT and ATC are different in part (a) because a unit receiving the treatment, usually means that $X$ was positive which means that $Y_i(1) - Y_i(0)$ will often be positive, while a unit not receiving the treatment usually means that $X$ was negative and thus $Y_i(1) - Y_i(0)$ will often be negative. Since the ATT is taken as an average over 1000 units, then the added noise of the normal distribution will essentially cancel itself out, and we are left with an ATT that is usually positive and an ATC that is usually negative.

The ATT and ATC are the same in part (b) because while a unit receiving the treatment usually means that $X$ was positive, $Y_i(1) - Y_i(0)$ is always just going to be 1 plus some noise that on average will cancel itself out. In the same way, a unit that did not receive the treatment usually means that $X$ was negative, but again $Y_i(1) - Y_i(0)$ will always just be 1 plus some noise that on average will be cancelled out.

The ATT and ATC are the same in part (c) because there is no confounding and the treatment is randomized. Thus, when we average over a large number of units we expect $Y(1) - Y(0)$ to be 0.

**e**

We can simply use the law of iterated expectation

$$E[Y(1) - Y(0)] = E\Big[E[Y(1) - Y(0) \mid D]\Big]$$
$$= E[Y(1) - Y(0) \mid D = 1]P(D = 1) + E[Y(1) - Y(0) \mid D = 0]P(D = 0)$$
$$= \text{ATT} \cdot P(D = 1) + \text{ATC} \cdot P(D = 0)$$

**f**

```
dgp_f <- function(n) {
  x <- runif(n, -1, 1)
  d <- rbinom(n , 1, 0.5 + 0.5*x)
  Y_0 = x  + rnorm(n, 0, 1)
  Y_1 = 2*x + rnorm(n, 0 , 1)
  tibble(
    x = x,
    d = d,
    Y_0 = Y_0,
    Y_1 = Y_1
  )
}

f_res <- run_sim(1000, dgp_f)

f_res %>%
  ggplot(aes(x = value, fill = var)) +
    geom_density()
```