hw3

1

```r
dgp_p1 <- function() {
  ### Define number of schools, and number of students in the school
  G <- 50 # number of schools
  ng <- 10 # number of students in each school
  n <- G*ng # total n-size

  ### Assign students to schools
  school <-rep(seq(1,G), ng)
  school <- sort(school)

  ### Gamma (school-varying intercept for PO's)
  gamma <- rnorm(G, sd=sqrt(2))[school]

  ### Generate potential outcomes
  y0 <- gamma + rnorm(n)
  y1 <- 2 + gamma + rnorm(n)

  ### Put everything into a data-frame
  data <- data.frame("school" = school, "y0" = y0, "y1" = y1)

  return(data)
}
```

a

There are 10 students per school and 50 schools total, meaning taht there are 500 total students. The ATE here is 2, and the factors that influence the potential outcomes are some random noise from the $\mathcal{N}(0,1)$ instances added to each one, as well as the school specific random intercept called 'gamma'.

b

1) Complete: Randomly draw 250 students to be treated
2) Bernoulli: Probability of treatment for each student is 0.5
3) Cluster: Randomly draw 25 schools to be treated
4) Statified: Within each school, randomly draw 5 students to be treated

```
library(tidyverse)

# empty results df to fill
res <- tibble(
  complete = rep(0, 1000),
  bernoulli = rep(0, 1000),
  cluster = rep(0, 1000),
  stratified = rep(0, 1000)
)

for (i in 1:1000) {
  data <- dgp_p1()

  # complete
  row_ids <- sample(1:500, 250)
  complete <- data %>%
    mutate(
      id = row_number(),
      D = if_else(id %in% row_ids, 1, 0)
      ) %>%
    select(-id)

  # bernoulli
  bernoulli <- data %>%
    mutate(D = rbinom(500, 1, 0.5))

  # cluster
  school_ids <- sample(1:50, 25)
  cluster <- data %>%
    group_by(school) %>%
    mutate(
      group_id = cur_group_id(),
      D = if_else(group_id %in% school_ids, 1 , 0)
    ) %>%
    ungroup() %>%
```

```
    select(-group_id)

  # stratified
  within_school_ids <- data %>%
    group_by(school) %>%
    mutate(chosen_id = row_number()) %>%
    slice_sample(n = 5) %>%
    ungroup() %>%
    select(school, chosen_id) %>%
    mutate(D = 1)

  stratified <- data %>%
    group_by(school) %>%
    mutate(
      id = row_number()
      ) %>%
    ungroup() %>%
    left_join(
      within_school_ids, by = c("id" = "chosen_id", "school")
      ) %>%
    mutate(D = replace_na(D, 0)) %>%
    select(-id)

  data_list <- list(complete, bernoulli, cluster, stratified)

  dim <- function(data){

    return(mean(data[data$D == 1, ]$y1) - mean(data[data$D == 0, ]$y0))

  }

  res[i, ] <- data_list %>%
    map(.f = dim)

}
```

```
res %>%
  pivot_longer(cols = everything(), names_to = "sampling_mechanism", values_to = "dim") %>
  ggplot(aes(x = dim, fill = sampling_mechanism)) +
  geom_density(alpha = 0.4, show.legend = F) +
  facet_wrap(~ sampling_mechanism, ncol = 1) +
```

```
theme_minimal()
```