hw5

1

a

Identify the ATE under conditional ignorability:

$$
\begin{aligned}
E[Y(1) - Y(0)] &= E\Big[E[Y(1) \mid X]\Big] - E\Big[E[Y(0) \mid X]\Big] \qquad \text{iterated expectation} \\
&= E\Big[E[Y(1) \mid X,\ D = 1]\Big] - E\Big[E[Y(0) \mid X,\ D = 0]\Big] \qquad \text{CI}
\end{aligned}
$$

b

2

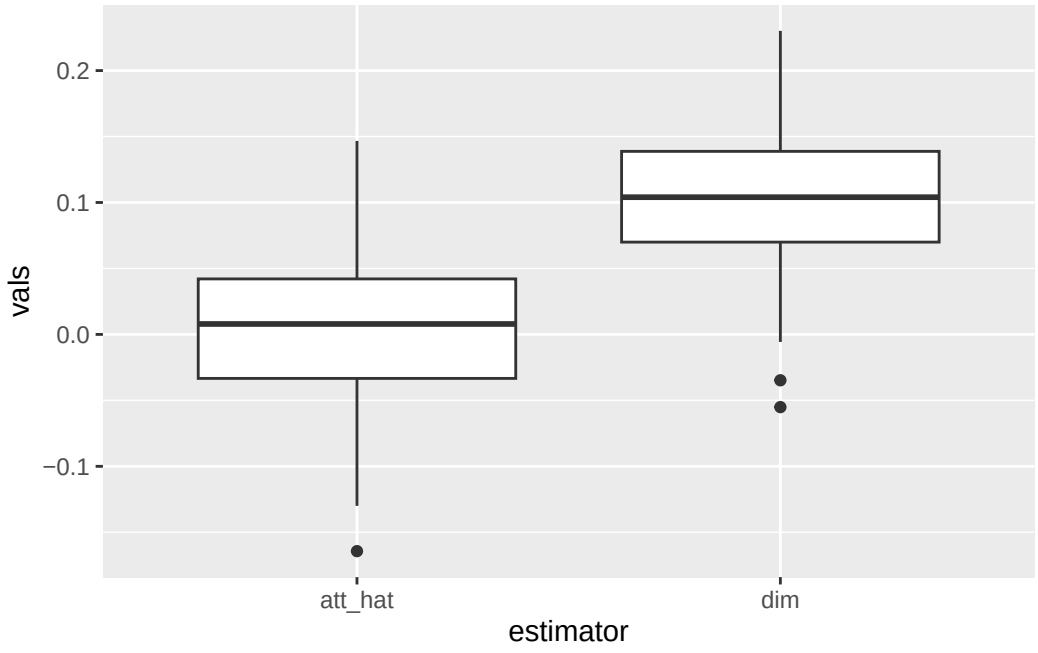
a

```
library(tidyverse)
dgp_p2 <- function(n) {
  tibble(
    x1 = rnorm(n, 0, 1),
    x2 = rchisq(n, 1),
    prob_D = exp(0.5*x1 + 0.5*x2 - 0.5)/(1 + exp(0.5*x1 + 0.5*x2 - 0.5)),
    D = rbinom(n, 1, prob_D),
    prob_Y = exp(0.6*x1 + 0.2*x2 + 0.5*x1*x2)/(1 + exp(0.6*x1 + 0.2*x2 + 0.5*x1*x2)),
    Y = rbinom(n, 1, prob_Y)
  ) %>%
    select(-prob_D, - prob_Y)
}
```

```r
att_est <- function(dat) {
  mod_y <- glm(Y ~ x1 + x2 + x1*x2, dat[dat$D == 0, ], family = "binomial")
  p1 <- mean(dat[dat$D == 1, "Y", drop = T])
  p2 <- mean(predict(mod_y, newdata = dat[dat$D==1, ], type = "response"))
  est_att <- p1 - p2

  return(est_att)
}

bootstrap <- function(dat){
  ests <- rep(0, 400)
  for(i in 1:400) {
    dat_boot <- dat %>%
      slice_sample(n = 400, replace = T)

    ests[i] <- att_est(dat_boot)
  }
  return(ests)
}


sim <- function(df) {
  dim <- lm(Y ~ D, df)$coefficients[2]
  est_att <- att_est(df)
  boots <- bootstrap(df)
  ci <- quantile(boots, c(0.025, 0.975))

  tibble(
    dim = dim,
    att_hat = est_att,
    lower = ci[1],
    upper = ci[2]
  )
}

data_list <- list()
for(i in 1:200){
  data_list[[i]] <- dgp_p2(400)
}

res <- data_list %>%
```

```
  map_dfr(sim)
```

```
library(here)

res <- read_csv(here("data", "hw5_res.csv"))

res %>%
  pivot_longer(cols = c(dim, att_hat), names_to = "estimator", values_to = "vals") %>%
  ggplot(aes(x = estimator, y = vals)) +
  geom_boxplot()
```




```
res %>%
  rowwise() %>%
  mutate(covers = between(0, lower, upper)) %>%
  ungroup() %>%
  summarise(coverage = mean(covers))
```

```
# A tibble: 1 x 1
  coverage
     <dbl>
1    0.945
```