

Homework 7

MA 590 Special Topics: Causal Inference

Aukkawut Ammartayakun

03 April, 2023

In May 2020 many US counties implemented a mask mandate to slow the spread of COVID-19.

The dataset `covidProcessed.csv` contains demographic data on US counties, along with covid infection and death totals and per-capita rates for April, 2020 (`aprilInfec`, `aprilDeaths`, `aprilInfecPC`, and `aprilDeathsPC`, respectively), covid death totals and per-capita rates for June, 2020 (`juneDeaths`, `juneDeathsPC`) and whether the county implemented a mask mandate in May 2020 (`Z`).

Problem 1

Use propensity score matching with this dataset to estimate the effect of a county's decision to impose a mask mandate (`Z`) on its per-capita covid death rate in June (`juneDeathsPC`). You may use all of the other variables (except `juneDeaths`) as covariates. Be sure to assess covariate balance before estimating effects. You may estimate the ATT, ATE, or another weighted average effect.

Solution

```
lm_robust(juneDeathsPC ~ Z, data = data)
```

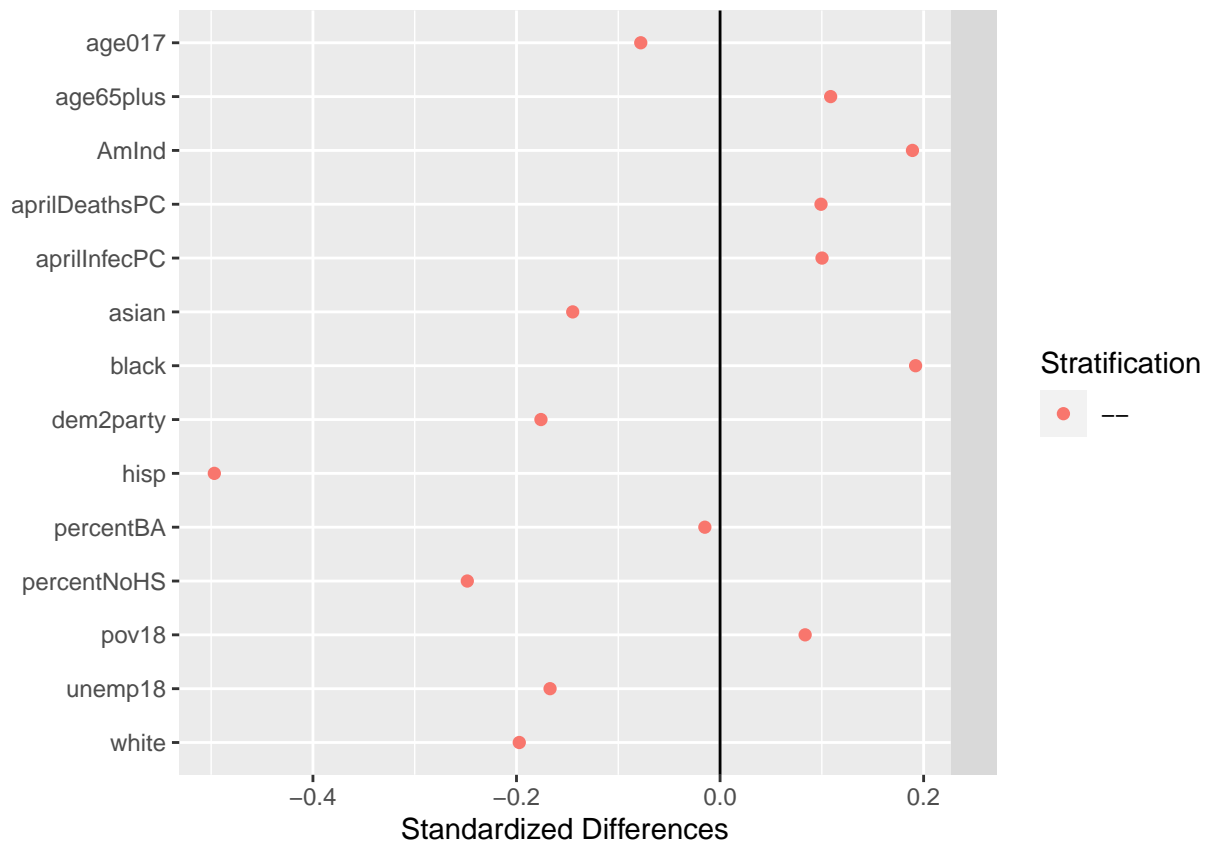
```
##               Estimate   Std. Error   t value   Pr(>|t|)   CI Lower
## (Intercept) 0.0049038498 0.0002960837 16.5623769 1.258836e-58 0.0043232637
## ZTRUE       0.0001075958 0.0003927827  0.2739321 7.841588e-01 -0.0006626061
##               CI Upper   DF
## (Intercept) 0.0054844360 2576
## ZTRUE       0.0008777977 2576
```

```
# covariate balance
(bal = balanceTest(Z ~ percentNoHS + percentBA + pov18 + unemp18 + age017 +
  age65plus + black + asian + hisp + white + AmInd + dem2party + aprilDeathsPC +
  aprilInfecPC, data = data))
```

```
##               strata():      --
##               stat      Treatment  Control adj.diff std.diff      z
## vars
## percentNoHS           13.3      15.0      -1.65    -0.25     -6.14    ***
## percentBA             21.4      21.6     -0.146    -0.01     -0.36
## pov18                 0.149      0.144     0.0046     0.08      2.01
## unemp18               4.18       4.41     -0.229    -0.17     -3.98    ***
## age017                0.217      0.219    -0.00258   -0.08     -1.91
## age65plus             0.194      0.189     0.00511    0.11      2.66     *
## black                 0.1201     0.0931     0.0269     0.19      4.36    ***
## asian                 0.0186     0.0240    -0.00545   -0.14     -3.74    **
```

```
## hisp          0.0765  0.1538  -0.0773  -0.50  -12.62  ***
## white         0.836   0.865   -0.0294  -0.20   -4.55   ***
## AmInd         0.0256  0.0173  0.00835  0.19    4.15   ***
## dem2party     0.312   0.341   -0.0291  -0.18   -4.20   ***
## aprilDeathsPC 0.001180 0.000892 0.000287 0.10    2.25   .
## aprilInfecPC  0.0279  0.0227  0.00516  0.10    2.34   .
## ---Overall Test---
##      chisquare df  p.value
## --      296 14 1.02e-54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(bal)
```



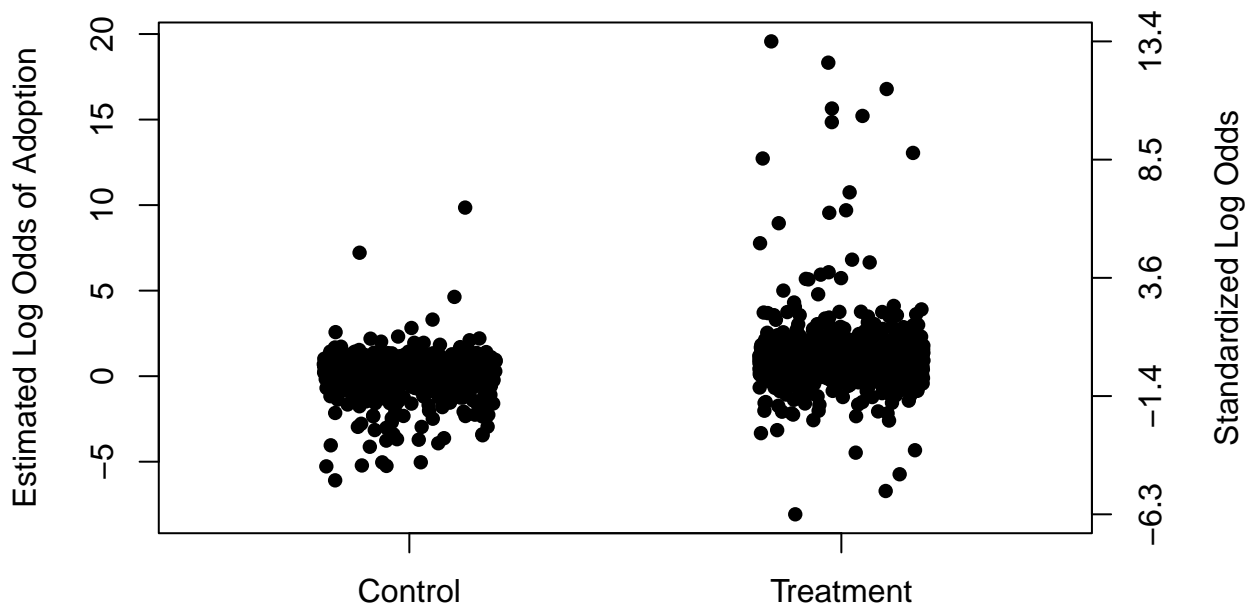
```
psMod1 <- glm(Z ~ percentNoHS + percentBA + pov18 + unemp18 + age017 +
  age65plus + black + asian + hisp + white + AmInd + dem2party + aprilDeathsPC +
  aprilInfecPC, data = data, family = binomial(logit))
summary(psMod1)
```

```
##
## Call:
## glm(formula = Z ~ percentNoHS + percentBA + pov18 + unemp18 +
##      age017 + age65plus + black + asian + hisp + white + AmInd +
##      dem2party + aprilDeathsPC + aprilInfecPC, family = binomial(logit),
##      data = data)
##
## Deviance Residuals:
```

##	Min	1Q	Median	3Q	Max
----	-----	----	--------	----	-----

```
## -4.4396 -1.2007 0.6694 0.8734 4.0188
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -12.194467  11.190988 -1.090  0.27586
## percentNoHS -0.058187  0.013886 -4.190 2.79e-05 ***
## percentBA   -0.005681  0.008490 -0.669  0.50344
## pov18        9.493098  1.501580  6.322 2.58e-10 ***
## unemp18     -0.369579  0.046629 -7.926 2.27e-15 ***
## age017       5.005530  2.120416  2.361  0.01824 *
## age65plus    1.681058  1.428130  1.177  0.23915
## black       14.204199  11.113090  1.278  0.20120
## asian       12.244992  12.383554  0.989  0.32276
## hisp       -4.253689  0.488247 -8.712 < 2e-16 ***
## white      12.646023  11.151250  1.134  0.25678
## AmInd      34.915898  11.821793  2.954  0.00314 **
## dem2party   -1.126211  0.402592 -2.797  0.00515 **
## aprilDeathsPC -10.872305 28.917354 -0.376  0.70693
## aprilInfecPC  4.543722  1.737155  2.616  0.00891 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 3344.0  on 2577  degrees of freedom
## Residual deviance: 2975.7  on 2563  degrees of freedom
## AIC: 3005.7
##
## Number of Fisher Scoring iterations: 6
```

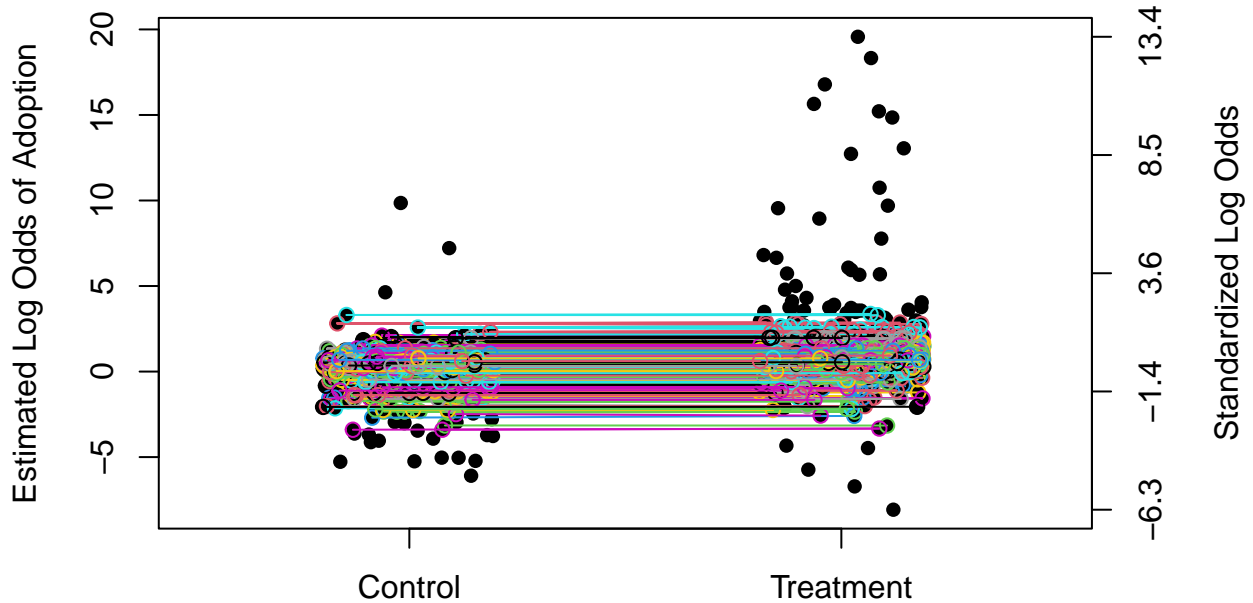
```
plotPS(psMod1)
```



```
dist = match_on(psMod1, caliper = 0.2, data = data)
fullCaliper <- fullmatch(dist, data = data)
summary(fullCaliper, psMod1)
```

```
## Structure of matched sets:
## 1:0 5+:1 4:1 3:1 2:1 1:1 1:2 1:3 1:4 1:5+ 0:1
## 56 90 26 45 77 229 62 33 20 20 23
## Effective Sample Size: 798
## (equivalent number of matched pairs).
##
## Balance test overall result:
## chisquare df p.value
## 32.3 14 0.00363
```

```
plotMatch(fullCaliper, mod = psMod1)
```

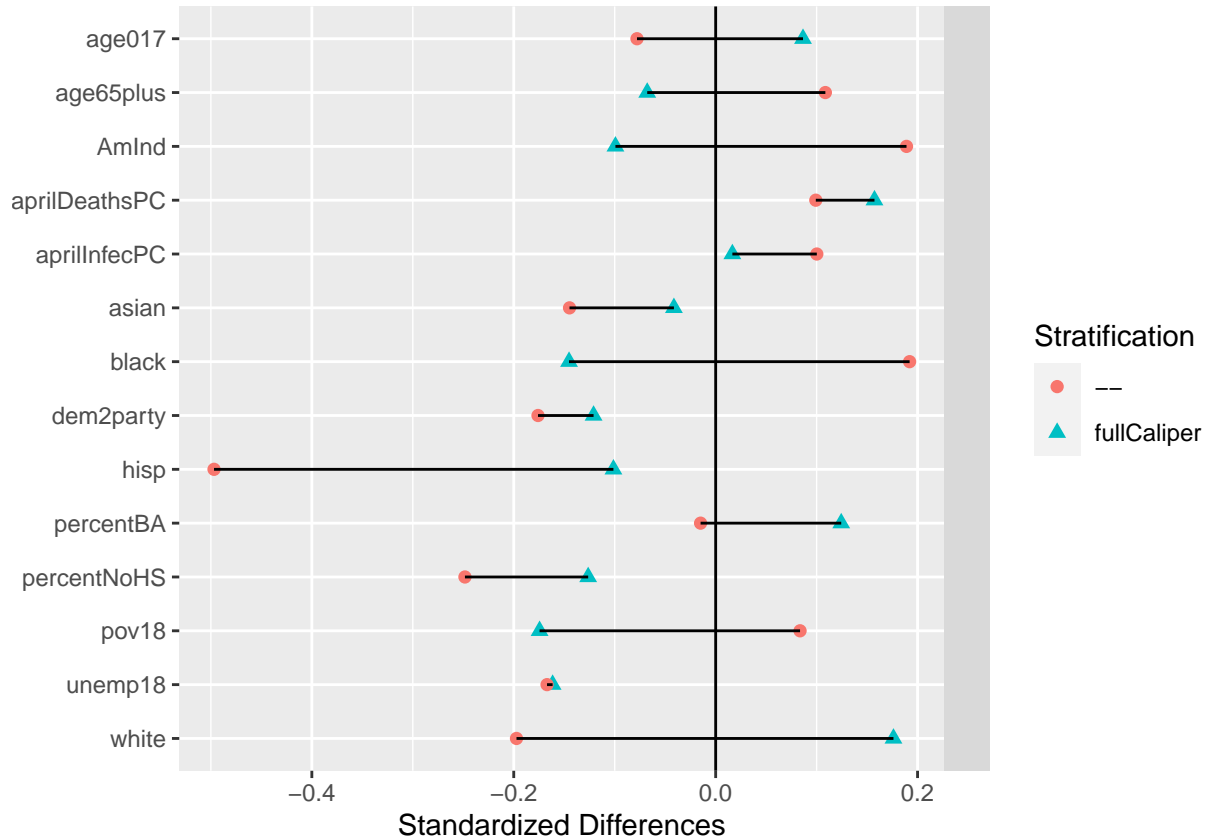


```
(bal = balanceTest(Z ~ percentNoHS + percentBA + pov18 + unemp18 + age017 +
  age65plus + black + asian + hisp + white + AmInd + dem2party + aprilDeathsPC +
  aprilInfecPC + strata(fullCaliper), data = data))
```

```
## strata(): fullCaliper
## stat Treatment Control adj.diff std.diff z Treatment Control
## vars
## percentNoHS 13.4 14.2 -0.838 -0.13 -1.71 13.3 15.0
## percentBA 21.5 20.3 1.213 0.12 1.63 21.4 21.6
## pov18 0.147 0.157 -0.0096 -0.17 -2.49 0.149 0.144
## unemp18 4.12 4.34 -0.221 -0.16 -0.17 4.18 4.41
## age017 0.216 0.213 0.00287 0.09 -2.53 0.217 0.219
## age65plus 0.194 0.197 -0.00319 -0.07 0.91 0.194 0.189
## black 0.1221 0.1425 -0.0204 -0.15 -2.47 0.1201 0.0931
## asian 0.0188 0.0204 -0.00156 -0.04 -0.14 0.0186 0.0240
## hisp 0.0748 0.0905 -0.0158 -0.10 -2.35 0.0765 0.1538
## white 0.841 0.815 0.0262 0.18 2.28 0.836 0.865
## AmInd 0.0184 0.0228 -0.00439 -0.10 1.54 0.0256 0.0173
## dem2party 0.311 0.331 -0.0200 -0.12 0.66 0.312 0.341
## aprilDeathsPC 0.001166 0.000710 0.000456 0.16 0.45 0.001180 0.000892
## aprilInfecPC 0.0276 0.0268 0.000853 0.02 0.43 0.0279 0.0227
## ---Overall Test---
## chisquare df p.value
## fullCaliper 32.3 14 3.63e-03
```

```
## --                295.5 14 1.02e-54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(bal)
```



```
# estimate the effect
lm_robust(juneDeathsPC ~ Z + predict(psMod1, type = "r"), data = data,
          fixed_effects = ~fullCaliper)
```

```
## Warning in eval(quote({: Some observations have missingness in the fixed_effects
## variable(s) but not in the outcome or covariates. These observations have been
## dropped.
```

```
##
##           Estimate Std. Error   t value Pr(>|t|)
## ZTRUE          -0.0006942413 0.000412008 -1.6850189 0.09214957
## predict(psMod1, type = "r") -0.1592063868 0.211319296 -0.7533926 0.45130764
##
##           CI Lower   CI Upper   DF
## ZTRUE          -0.001502278 0.0001137957 1895
## predict(psMod1, type = "r") -0.573649304 0.2552365308 1895
```

Problem 2

The IPW estimator for the ATT is

$$\hat{\tau}_{IPW}^{ATT} = \frac{1}{n_1} \sum_i Y_i \left(Z_i - \frac{e(X_i)}{1 - e(X_i)} (1 - Z_i) \right)$$

where $e(X_i)$ is subject i 's propensity score.

- a) Show that if strong ignorability holds and propensity scores are known exactly—i.e. if $e(X_i) = \Pr(Z_i = 1)$ —then

$$\mathbb{E}[\hat{\tau}_{\text{IPW}}^{\text{ATT}}] = \frac{1}{n_1} \sum_i e(X_i) \tau_i = \mathbb{E}[\tau | Z = 1]$$

- b) Use the propensity scores from (1) and IPW to estimate the effect of mask mandates on June 2020 covid death rates for counties that adopted them.

Solution

a)

We want to show that if strong ignorability holds and propensity scores are known exactly, then

$$\mathbb{E}[\hat{\tau}_{\text{IPW}}^{\text{ATT}}] = \frac{1}{n_1} \sum_i e(X_i) \tau_i = \mathbb{E}[\tau | Z = 1]$$

Let Y_i be the outcome of interest, Z_i be the treatment, X_i be the covariates, and $e(X_i)$ be the propensity score. We have

$$\mathbb{E}[\hat{\tau}_{\text{IPW}}^{\text{ATT}}] = \frac{1}{n_1} \mathbb{E} \left[\sum_i Y_i Z_i - Y_i \frac{e(X_i)}{1 - e(X_i)} (1 - Z_i) \right]$$

Linearize the expectation, we will find that the first term $\mathbb{E}[Y_i Z_i]$ is $n_1 e(X_i) \mathbb{E}[y(1)]$ as shown in the lecture.

Recall that ATT is defined as

$$\tau = \frac{1}{n_1} \sum_i y_i(1) - y_i(0)$$

In the similar manner, we can show that the second term, $\mathbb{E} \left[Y_i \frac{e(X_i)}{1 - e(X_i)} (1 - Z_i) \right]$ is $n_1 e(X_i) \mathbb{E}[y(0)]$ by letting $Y_i = y_i(1)Z_i + y_i(0)(1 - Z_i)$ and $\mathbb{E}[1 - Z_i] = 1 - e(X_i)$. Thus, the result is now as desired.

b)

```
# ipw
obs = data %>%
  mutate(match = fullCaliper) %>%
  group_by(match) %>%
  mutate(n1 = sum(Z), n0 = sum(1 - Z), wols = n1 * n0 / (n1 + n0), w = n1 / wols) %>%
  ungroup()

lm_robust(juneDeathsPC ~ Z + predict(psMod1, type = "r"), data = obs, fixed_effects = ~match,
  weights = w)
```

```
## Warning in eval(quote({: Some observations have missingness in the fixed_effects
## variable(s) but not in the outcome or covariates. These observations have been
## dropped.
```

```
##               Estimate   Std. Error   t value   Pr(>|t|)
## ZTRUE              0.0003748326 0.0005417699  0.6918668 0.4891057
## predict(psMod1, type = "r") -0.3107092364 0.2704484386 -1.1488668 0.2507559
##               CI Lower   CI Upper   DF
## ZTRUE              -0.0006876955 0.001437361 1895
## predict(psMod1, type = "r") -0.8411172114 0.219698739 1895
```

Problem 3

Do you believe these estimates? Why or why not? Is there an observational design (perhaps with more/different observational data) that you would find more convincing?

Solution

Yes, because the result suggests that wearing mask actually negatively reduce the death rate. The longitudinal data might help address the problem as there are more data point in time to compare rather than the instatenous difference.