Info 371 Lab 4

Yasmine Hejazi

1 Was Montana Meth Prevention Effective?

In 2005-2007, Montana conducted an aggressive ad campaign to decrease the meth abuse by the youth. We will analyze the efficacy of the campaign using Youth Risk Behavior Survey (YRBS) data. We will do this by implementing: a) cross-sectional estimator; b) before- after estimator; c) differences-in-differences (DiD) estimator. The full data and documentation can be obtained from https://www.cdc.gov/healthyyouth/data/yrbs/data.htm.

```
yrbs <- read.delim("yrbs.tsv")
yrbs$before <- yrbs$year < 2006
```

2 Before-After Estimator

If the campaign had any effect, it should have decreased the amphetamine use in Montana. Let's compare the meth use (mean value) in Montana before and after the campaign. In each case, let's interpret the most important outcomes, and compare the results for all three approaches.

First, let's present it as a simple table.

before	avg_meth_use
FALSE	0.0396259
TRUE	0.0631416

Next, let's do it using regression without any other controls (no age, sex, tv), focusing on Montana.

```
##
## Call:
## lm(formula = meth ~ before, data = subset(yrbs, state == "MT"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07673 -0.07673 -0.03811 -0.03811  0.96189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.038115   0.003319  11.484   <2e-16 ***
## beforeTRUE  0.038619   0.004695   8.225   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2318 on 9752 degrees of freedom
## Multiple R-squared:  0.00689,    Adjusted R-squared:  0.006788
## F-statistic: 67.66 on 1 and 9752 DF,  p-value: < 2.2e-16
```

Next, let's do it using regression *with* other controls (no age, sex, tv), focusing on Montana.

```
##
## Call:
```

```
## lm(formula = meth ~ before + age + sex + tv, data = subset(yrbs,
##     state == "MT"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09527 -0.07691 -0.05361 -0.03305  0.99713
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.147630   0.037744  -3.911 9.24e-05 ***
## beforeTRUE   0.038590   0.004694   8.221 2.28e-16 ***
## age          0.011814   0.002385   4.953 7.44e-07 ***
## sexM        -0.003071   0.004703  -0.653    0.514
## tv           0.003478   0.004707   0.739    0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2316 on 9749 degrees of freedom
## Multiple R-squared:  0.00943,    Adjusted R-squared:  0.009024
## F-statistic:  23.2 on 4 and 9749 DF,  p-value: < 2.2e-16
```

In the first approach, we can easily identify through means that the average meth use decreased from before program to after program by **0.02351**. The second approach gives us more insight to how significant the *before* variable is. A t value of 8.225 shows that this is very significant. After the program has a lower estimate of meth usage than before the program by **0.038619**. Finally, we can introduce other controls to gain even more insight in factors that affect meth usage. We see similarly that after the program has a lower estimate than before by **0.038590**, and another significant variable is age, with a t value of 4.953.

3 Cross-Sectional Estimator

Alternatively, if the project had an effect, Montana should have lower rate of meth use afterwards. Now compare the meth use (mean value) in Montana and nationally after the project. Let's interpret the most important outcomes, and compare the results for all three approaches.

Do it as a simple table: compare the average *post-project* meth use in MT and nationally.

state	avg_meth_use
MT	0.0381148
XX	0.0399227

Do it using regression without any other controls.

```
##
## Call:
## lm(formula = meth ~ state, data = yrbs %>% filter(before == FALSE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.03992 -0.03992 -0.03992 -0.03992  0.96189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.038115   0.002793  13.648   <2e-16 ***
```

```
## stateXX     0.001808   0.003055   0.592    0.554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1951 on 29726 degrees of freedom
## Multiple R-squared:  1.179e-05,  Adjusted R-squared:  -2.185e-05
## F-statistic: 0.3503 on 1 and 29726 DF,  p-value: 0.5539
```

Run a regression with full controls.

```
##
## Call:
## lm(formula = meth ~ state + age + sex + tv, data = yrbs %>% filter(before ==
##     FALSE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05267 -0.04474 -0.03838 -0.03203  0.98240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.064948   0.017965  -3.615 0.000301 ***
## stateXX      0.001516   0.003071   0.494 0.621500
## age          0.006355   0.001125   5.648 1.64e-08 ***
## sexM         0.008073   0.002266   3.563 0.000367 ***
## tv          -0.001576   0.002306  -0.683 0.494342
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1949 on 29723 degrees of freedom
## Multiple R-squared:  0.001558,   Adjusted R-squared:  0.001424
## F-statistic:  11.6 on 4 and 29723 DF,  p-value: 2.073e-09
```

In the first approach, we can easily identify through means that the average meth use for Montana is less than the national level by **0.00181**. The second approach gives us more insight to how significant the *state* variable is. A t value of 0.592 shows that *state* is not very significant. Montana has a lower estimate of meth usage than National by **0.00181**, very similar to the first approach. Finally, we can introduce other controls to gain even more insight in factors that affect meth usage by state. We see similarly that Montana has a lower estimate than National by **0.001516** , but again is not significant. We see more significant factors of age (t=5.648) and sex (t=3.563) appear.

4 Differences-in-Differences Estimator

However, both the long-time trends and levels in meth use in Montana and elsewhere may be different. Now let's see how much more will meth use decline in Montana compared to the national average.Compare the differences in trends in meth use in Montana and nationally between pre-project and post-project years. In each case interpret the most important outcomes, and compare the results for all three approaches.

Do it as a simple table: report average use before and after the campaign in both Montana and nationally. Compute the effect by calculating the trend difference.

state	before	avg_meth_use
MT	FALSE	0.0381148
MT	TRUE	0.0767337
XX	FALSE	0.0399227

state	before	avg_meth_use
XX	TRUE	0.0603195

Do it using regression without any other controls.

```
##
## Call:
## lm(formula = meth ~ state * before, data = yrbs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07673 -0.06032 -0.03992 -0.03992  0.96189
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.038115   0.003147  12.110  < 2e-16 ***
## stateXX             0.001808   0.003443   0.525 0.599453
## beforeTRUE          0.038619   0.004452   8.674  < 2e-16 ***
## stateXX:beforeTRUE -0.018222   0.004881  -3.733 0.000189 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2199 on 58073 degrees of freedom
## Multiple R-squared:  0.00324,    Adjusted R-squared:  0.003189
## F-statistic: 62.92 on 3 and 58073 DF,  p-value: < 2.2e-16
```

Run a regression with full controls

```
##
## Call:
## lm(formula = meth ~ state * before * age * sex * tv, data = yrbs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10630 -0.06035 -0.04941 -0.03551  0.99346
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -0.2133799  0.0900769  -2.369  0.01785 *
## stateXX                       0.1653008  0.1013536   1.631  0.10291
## beforeTRUE                    0.1337162  0.1322241   1.011  0.31188
## age                           0.0158518  0.0057207   2.771  0.00559 **
## sexM                          0.0600958  0.1359186   0.442  0.65839
## tv                            0.0131585  0.1375868   0.096  0.92381
## stateXX:beforeTRUE           -0.1126374  0.1499048  -0.751  0.45242
## stateXX:age                  -0.0103399  0.0064293  -1.608  0.10779
## beforeTRUE:age               -0.0059079  0.0084000  -0.703  0.48186
## stateXX:sexM                 -0.1090394  0.1529891  -0.713  0.47602
## beforeTRUE:sexM              -0.0140420  0.2000371  -0.070  0.94404
## age:sexM                     -0.0035574  0.0086084  -0.413  0.67943
## stateXX:tv                    0.0304703  0.1505327   0.202  0.83959
## beforeTRUE:tv                -0.1147261  0.1958342  -0.586  0.55799
## age:tv                       -0.0005417  0.0087637  -0.062  0.95072
## sexM:tv                       0.0975250  0.2005165   0.486  0.62671
```

```
## stateXX:beforeTRUE:age          0.0060629  0.0095062   0.638  0.52362
## stateXX:beforeTRUE:sexM        -0.0193087  0.2264864  -0.085  0.93206
## stateXX:age:sexM                0.0069007  0.0096791   0.713  0.47588
## beforeTRUE:age:sexM             0.0002807  0.0126758   0.022  0.98234
## stateXX:beforeTRUE:tv           0.0654778  0.2154598   0.304  0.76121
## stateXX:age:tv                 -0.0026199  0.0095765  -0.274  0.78441
## beforeTRUE:age:tv               0.0075115  0.0124773   0.602  0.54717
## stateXX:sexM:tv                -0.0902224  0.2198147  -0.410  0.68148
## beforeTRUE:sexM:tv             -0.0524176  0.2866941  -0.183  0.85493
## age:sexM:tv                    -0.0066747  0.0127380  -0.524  0.60028
## stateXX:beforeTRUE:age:sexM     0.0020715  0.0143270   0.145  0.88504
## stateXX:beforeTRUE:age:tv      -0.0045479  0.0137058  -0.332  0.74002
## stateXX:beforeTRUE:sexM:tv      0.1083337  0.3159079   0.343  0.73165
## stateXX:age:sexM:tv             0.0067921  0.0139484   0.487  0.62630
## beforeTRUE:age:sexM:tv          0.0035777  0.0182152   0.196  0.84429
## stateXX:beforeTRUE:age:sexM:tv -0.0077907  0.0200413  -0.389  0.69748
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2197 on 58045 degrees of freedom
## Multiple R-squared:  0.005067,   Adjusted R-squared:  0.004536
## F-statistic: 9.537 on 31 and 58045 DF,  p-value: < 2.2e-16
```

In the first approach, we see that in Montana, average meth use moves from **0.0767337** to **0.0381148** for before and after program time respectively, while in National, the average meth use moves from **0.0603195** to **0.0399227**. We see a much larger difference in Montana's change, where the average meth use started higher than National before program and ended less than National after program. The second approach gives us more insight to how significant the *state* and *before* variables are when crossed. *state* alone isn't significant, but *before* (t=8.674) and *before x state* (t=-3.733) shows a great significance in meth usage.