# Homework 6

## MA 590 Special Topics: Causal Inference

### Aukkawut Ammartayakun

### 02 March, 2023

Take a look at `Ludwig & Miller 2007.pdf`. You don't need to read all 50 pages–for our purposes sections II, V, and VII are the most important.

# Problem 1

Answer Sue Dynarski's 3 questions, in relation to Ludwig & Miller (2007): What is the causal question? What is the ideal experiment? What is the identification strategy?

## Solution

- The causal question is whether Head Start program has an effect on child mortality.
- The ideal experiment is to randomly assign children to Head Start program and non-Head Start program.
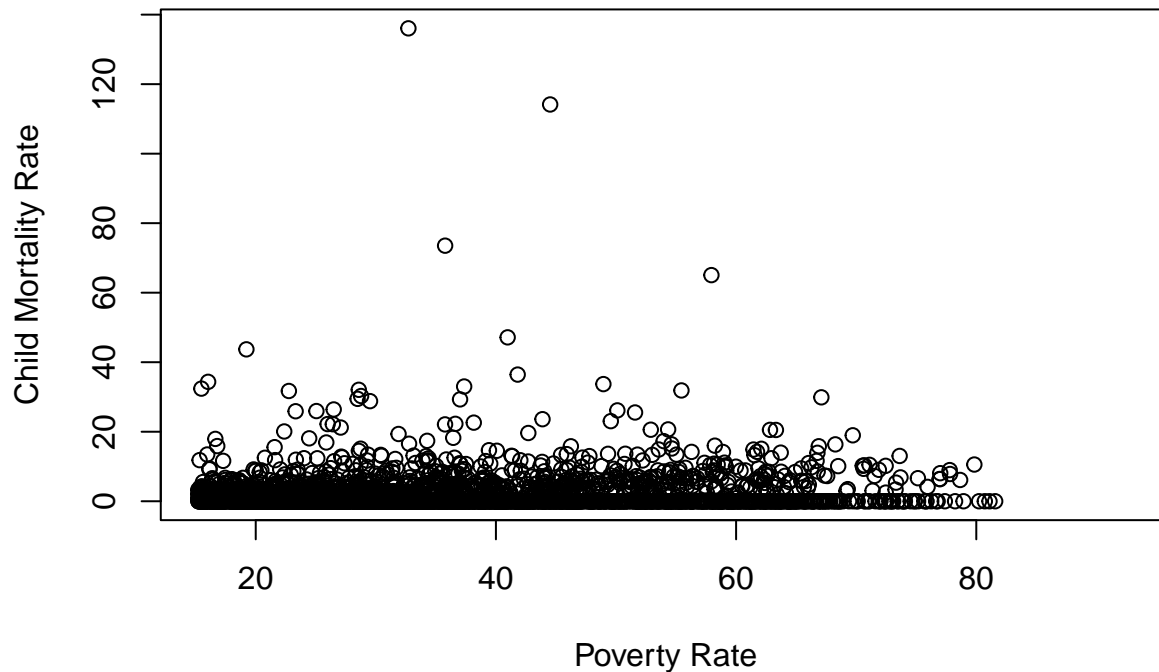- The identification strategy is to use the difference-in-difference method.

# Problem 2

Following their general approach, use the dataset `headstart.csv`, (more-or-less) replicate their results for Post-Head Start child mortality ages 5-9 (`mort_age59_related_postHS` in the dataset). Don't worry about getting exactly the same answer or model specification as in the paper, but if it's in an entirely different ballpark then maybe something went wrong.

## Solution

```
#scatter plot of poverty rate against child mortality rate
plot(data$povrate60, data$mort_age59_related_postHS, xlab = "Poverty Rate", ylab = "Child Mortality Rate
```

## Scatter Plot of Poverty Rate against Child Mortality Rate



```
#now, on our data
#cutoff at poverty rate >59.1984
data$Z=ifelse(data$povrate60>59.1984,1,0)
#bin the data by poverty rate, every 1%
data$bin = round(data$povrate60)
data$Z=as.factor(data$Z)
summary(data)
```

```
##    povrate60     mort_age59_related_postHS census1960_pop   census1960_pctsch1417 census1960_pctsch5
## Min.   :15.21   Min.   :  0.000           Min.   :    224   Min.   :  0.00       Min.   :0.0000
## 1st Qu.:24.15   1st Qu.:  0.000           1st Qu.:   9133   1st Qu.: 82.40       1st Qu.:0.5161
## Median :33.61   Median :  0.000           Median :  17126   Median : 87.30       Median :0.5524
## Mean   :36.79   Mean   :  2.253           Mean   :  38964   Mean   : 84.07       Mean   :0.5511
## 3rd Qu.:47.42   3rd Qu.:  2.824           3rd Qu.:  33417   3rd Qu.: 91.60       3rd Qu.:0.5891
## Max.   :93.07   Max.   :136.054           Max.   :2664438   Max.   :100.00       Max.   :0.7516
## NA's   :5       NA's   :24                NA's   :5         NA's   :27           NA's   :26
## census1960_pop534 census1960_pop25plus census1960_pcturban census1960_pctblack   Z            bin
## Min.   :     70   Min.   :    120       Min.   :  0.00      Min.   : 0.00       0  :2504   Min.   :
## 1st Qu.:   4000   1st Qu.:   5067       1st Qu.:  0.00      1st Qu.: 0.10       1  : 300   1st Qu.:
## Median :   7872   Median :   9249       Median : 28.30      Median : 1.50       NA's:   5  Median :
## Mean   :  18295   Mean   :  21712       Mean   : 29.15      Mean   :10.63                  Mean   :
## 3rd Qu.:  15521   3rd Qu.:  17714       3rd Qu.: 47.90      3rd Qu.:15.00                  3rd Qu.:
## Max.   :1180784   Max.   :1611158       Max.   :100.00      Max.   :83.40                  Max.   :
## NA's   :26        NA's   :26            NA's   :21          NA's   :21                     NA's   :
```

```
data%>%
mutate(Z=as.factor(Z))%>%
filter(!is.na(bin))%>%
group_by(bin,Z)%>%
summarize(AvgMortality=mean(mort_age59_related_postHS,na.rm=TRUE),n=n())%>%
```
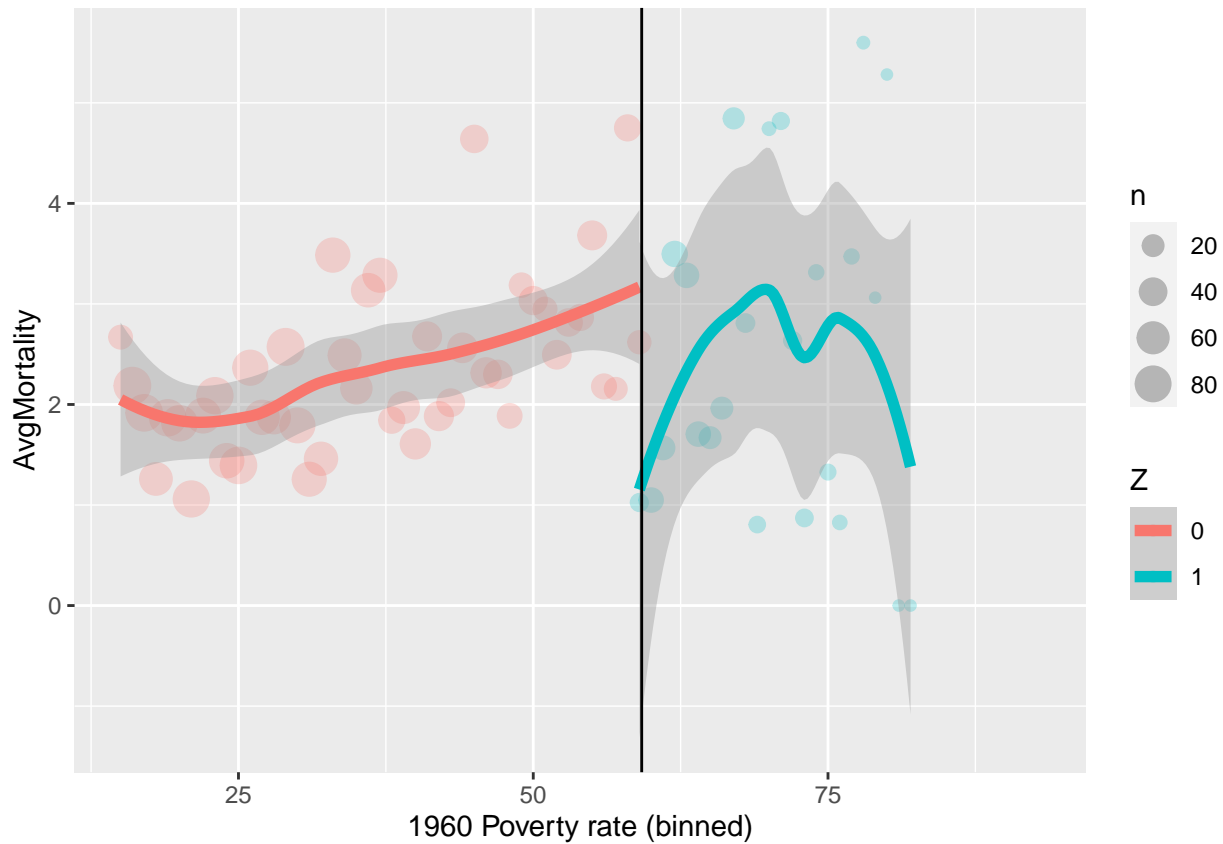
```
ggplot(aes(bin,AvgMortality, color=Z))+geom_point(aes(size=n),alpha=0.25)+geom_smooth(linewidth=2)+
geom_vline(xintercept=59.1984)+xlab("1960 Poverty rate (binned)")
```

```
## `summarise()` has grouped output by 'bin'. You can override using the `.groups` argument.
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



```
#library(rdrobust)
#rd = rdrobust(data$mort_age59_related_postHS, data$Z)
```

# Problem 3

Critically evaluate the result–what does it mean? Do you believe it? Why or why not?

**Solution**