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 regression discontinuity design (RDD) to estimate the effect of Head Start program on child mortality.

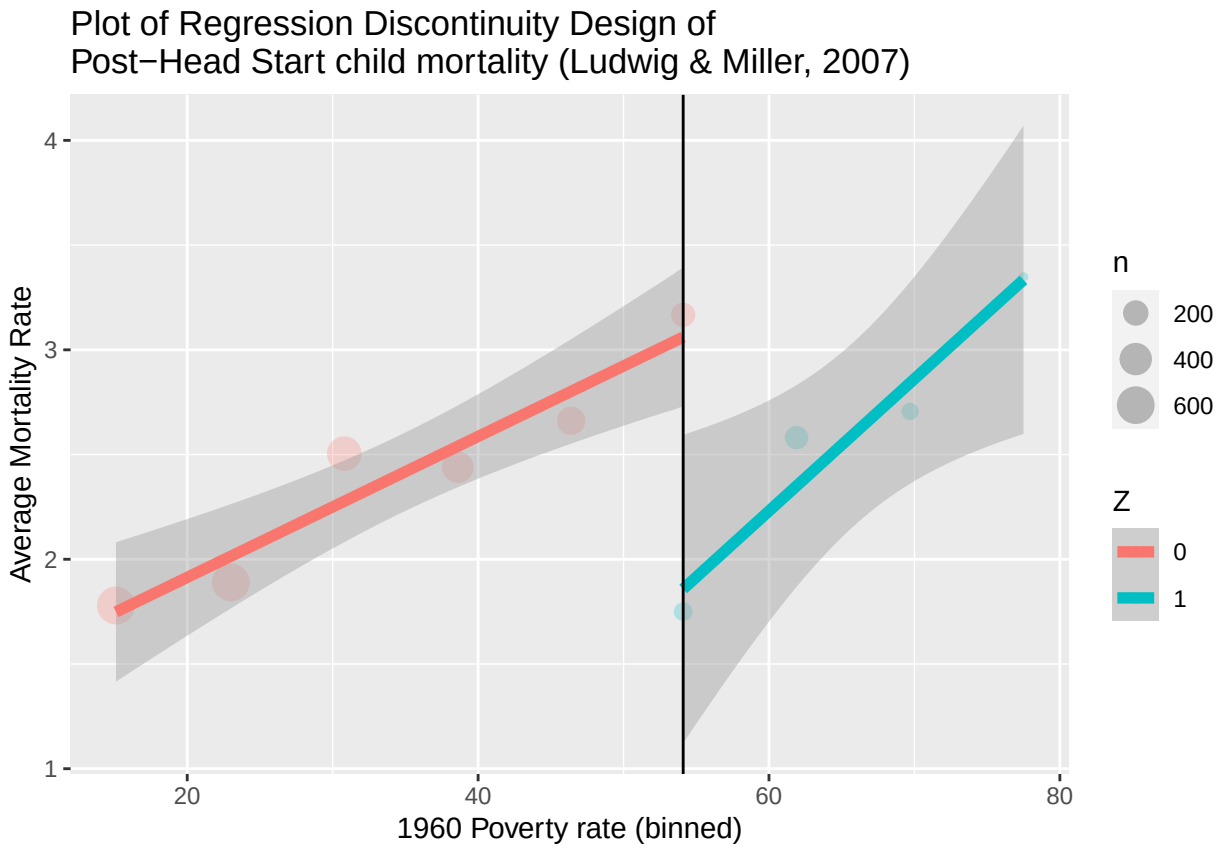
Problem 2

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

```
#cutoff at poverty rate >59.1984
data$Z=ifelse(data$povrate60>59.1984,1,0)
#paper uses 10 bins, defined by cut(data$povrate60,10)
bins = c(15.1,23,30.8,38.6,46.4,54.1,61.9,69.7,77.5,100) #UB of each bin
for (i in 1:length(bins)){
  data$bin[data$povrate60>bins[i]]=bins[i]
}
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, method = lm)+
geom_vline(xintercept=54.1)+xlab("1960 Poverty rate (binned)")+ #change due to binning
```

```
ggtitle("Plot of Regression Discontinuity Design of\nPost-Head Start child mortality (Ludwig & Miller, 2
ylab("Average Mortality Rate")
```




```
#using the model in the paper
#Yc = b0  + b1(Pc - P300) + a*Gc+ b2Gc(Pc -P300)+ vc
data$povrate_diff = data$povrate60-59.1984
summary(lm(mort_age59_related_postHS~povrate_diff + Z + Z*povrate_diff,data=data))
```

```
##
## Call:
## lm(formula = mort_age59_related_postHS ~ povrate_diff + Z + Z *
##     povrate_diff, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -3.240  -2.342  -1.903   0.753 133.841
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.127671   0.271555  11.518  < 2e-16 ***
## povrate_diff    0.034501   0.009505   3.630 0.000289 ***
## Z              -1.053547   0.603514  -1.746 0.080976 .
## povrate_diff:Z  0.017596   0.064219   0.274 0.784108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.714 on 2779 degrees of freedom
```

```
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.00506,    Adjusted R-squared:  0.003986
## F-statistic: 4.711 on 3 and 2779 DF,  p-value: 0.002773
```

Problem 3

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

Solution

The results shows that, first, there is the discontinuity in the average mortality rate given different poverty rate. Second, the estimate here are closed to the paper, that is the effect of headstart program on child mortality is in a negative way. Intuitively, this should not happen, because the headstart program is designed to help children in poverty. That means, there might be some confounding factors that affect the result. For example, the poverty rate might be correlated with other factors that affect the child mortality rate at this age range.