HW7

12351235

2025-02-26

```
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
library(estimatr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::recode() masks car::recode()
## x purrr::some()   masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
#Question 4

maimonides <- read_csv("maimonides.csv")
```

```
## Rows: 1829 Columns: 51
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): townname
## dbl (50): schlcode, c_size, c_boys, c_girls, c_numcl, c_pik, c_status, c_leo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#a
model_no_control <- lm_robust(avgverb ~ classize, maimonides)
model_control <- lm_robust(avgverb ~ classize + tipuach, maimonides)
summary(model_no_control)
```

```
##
## Call:
## lm_robust(formula = avgverb ~ classize, data = maimonides)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  67.7413    1.03395  65.517 0.000e+00 65.71345  69.7691 1824
## classize      0.1507    0.03168   4.756 2.129e-06  0.08853   0.2128 1824
##
## Multiple R-squared:  0.0143 ,    Adjusted R-squared:  0.01376
## F-statistic: 22.62 on 1 and 1824 DF,  p-value: 2.129e-06
```

```r
summary(model_control)
```

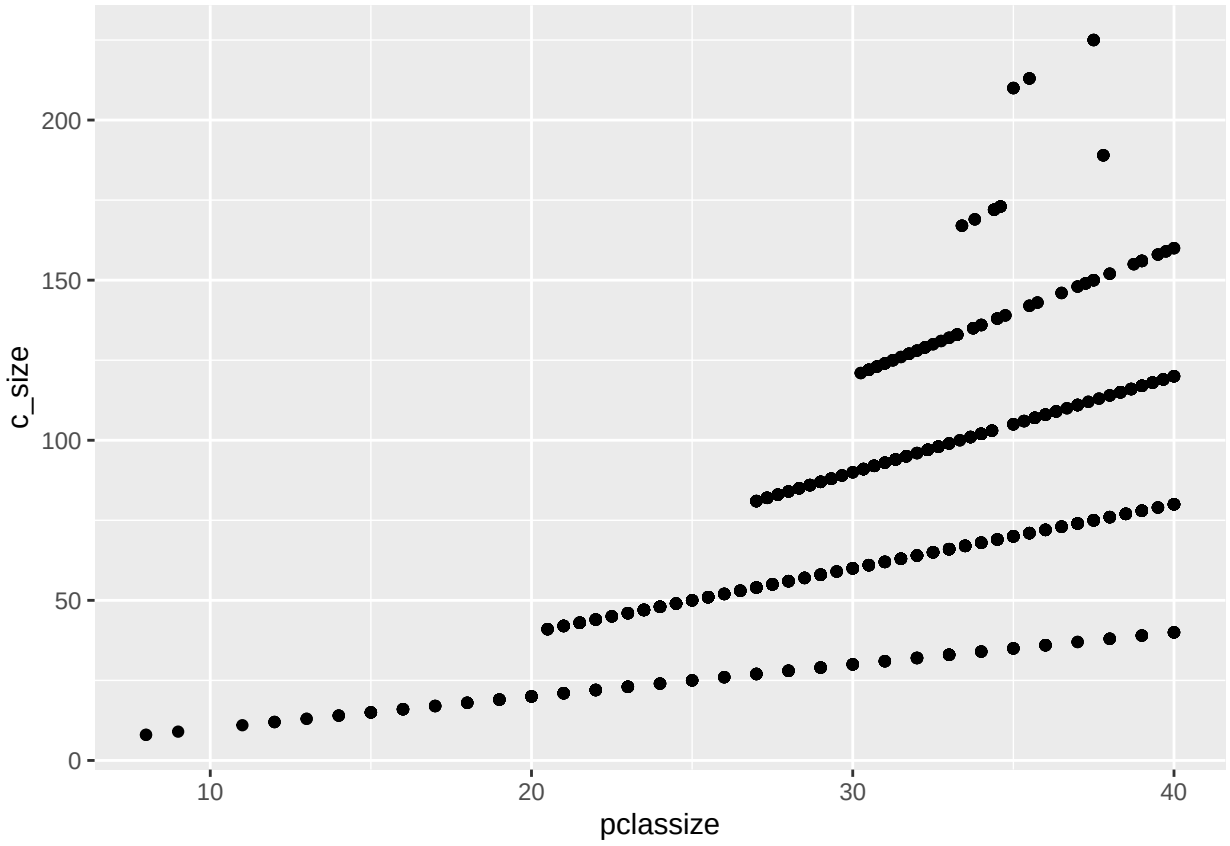
```
##
## Call:
## lm_robust(formula = avgverb ~ classize + tipuach, data = maimonides)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
## (Intercept) 78.05023    0.86173  90.574  0.000e+00 76.36014 79.74031 1823
## classize    -0.03681    0.02536  -1.451  1.469e-01 -0.08654  0.01293 1823
## tipuach     -0.32598    0.01380 -23.628 6.786e-108 -0.35304 -0.29892 1823
##
## Multiple R-squared:  0.2957 ,    Adjusted R-squared:  0.2949
## F-statistic:   284 on 2 and 1823 DF,  p-value: < 2.2e-16
```

In the model without the control variable, every additional student in a class is associated with a .15 point increase in average test scores, which is statistically and econonomically significant. In the model with the

control variable, every additional student is associated with a .04 decrease in average test score, which is not statistically or economically significant. We cannot interpret these results as causal because there are likely to be other variables that have an important effect on test scores that we are not accounting for. We can say that leaving out tipuach biases classize upwards.

```
#b
maimonides_pclassize <- maimonides |>
  mutate(pclassize = c_size/(floor((c_size - 1)/40) + 1))

ggplot(data = maimonides_pclassize,
       mapping = aes(x = pclassize,
                     y = c_size)) +
  geom_point()
```



The scatter plot shows class size going up in steps as school size increases. The plot shows that there is a relationship between class size and school size, so cov(classsize, school size) =! 0. Thus, class size does meet the exclusion restriction for an instrumental variable.

```
#c
first_stage_model <- lm_robust(classize ~ pclassize + tipuach, maimonides_pclassize)
summary(first_stage_model)
```

```
##
## Call:
## lm_robust(formula = classize ~ pclassize + tipuach, data = maimonides_pclassize)
##
```

```
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  7.15486    0.60405  11.845  3.083e-31  5.97017  8.33956 1826
## pclassize    0.77120    0.01831  42.129 1.500e-271  0.73530  0.80711 1826
## tipuach     -0.05282    0.00816  -6.473  1.230e-10 -0.06882 -0.03682 1826
##
## Multiple R-squared:  0.5479 ,    Adjusted R-squared:  0.5474
## F-statistic:  1187 on 2 and 1826 DF,  p-value: < 2.2e-16
```

Pclassize fits the instrumental because predicted class size is unlikely to have any effect on average test score except for its effect on actual class size. The fact that the first stage regression shows such a strong relationship between pclassize and class size provides evidence in support of this. The estimated coefficient is less than one because there are other factors that effect class size besides pclassize. In this case, the percentage of disadvantage students in a class also effects class size.

```
#d
iv_model <- ivreg(avgverb ~ classize + tipuach | pclassize + tipuach, data = maimonides_pclassize)
summary(iv_model)
```

```
##
## Call:
## ivreg(formula = avgverb ~ classize + tipuach | pclassize + tipuach,
##     data = maimonides_pclassize)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.2522  -3.8951   0.4418   4.4860  23.7823
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 79.16564    1.18468  66.824   <2e-16 ***
## classize    -0.07135    0.03601  -1.981   0.0477 *
## tipuach     -0.33036    0.01250 -26.431   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.693 on 1823 degrees of freedom
## Multiple R-Squared: 0.295,   Adjusted R-squared: 0.2942
## Wald test: 383.3 on 2 and 1823 DF,  p-value: < 2.2e-16
```

We can interpret the coefficient on classize as telling us that a one person increase in class size is associated with a -.07 point decrease in average test score.