Problem Set 4

Yiyun (Leo) Yao - yy3959 - (Recitation) 002

Due Dec 13, 2023

This homework must be turned in on Brightspace by Dec. 13 2023. It must be your own work, and your own work only – you must not copy anyone's work, or allow anyone to copy yours. This extends to writing code. You may consult with others, but when you write up, you must do so alone.

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. **No zip files will be accepted. Make sure we can read each line of code in the pdf document.** You should submit the following:

1. A compiled PDF file named yourNetID_solutions.pdf containing your solutions to the problems.
2. A .Rmd file containing the code and text used to produce your compiled pdf named yourNetID_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part
2. Do not include written answers as code comments.
3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

Problem 1 (100 points)

Despite the heated political and media rhetoric, there are a few causal estimates of the effect of expanded health insurance on healthcare outcomes. One landmark study, the Oregon Health Insurance Experiment, covered new ground by utilizing a randomized control trial implemented by the state of Oregon. To allocate a limited number of eligible coverage slots for the state's Medicaid expansion, about 30,000 low-income, uninsured adults (out of about 90,000 wait-list applicants) were randomly selected by lottery to be allowed to apply for Medicaid coverage. Researchers collected observable measure of health (blood pressure, cholesterol, blood sugar levels, and depression), as well as hospital visitations and healthcare expenses for 6,387 selected adults and 5,842 not selected adults.

For this problem, we will use the OHIE.dta file.

- treatment - selected in the lottery to sign up for Medicaid (instrument)
- ohp_all_ever_admin - Ever enrolled in Medicaid after notification of lottery results (compliance)
- tab2bp_hyper - Outcome: Binary indicator for elevated blood pressure (1 indicates a high blood pressure)
- tab2phqtot_high - Outcome: Binary indicator for depression
- tab4_catastrophic_exp_inp - Outcome: Indicator for catastrophic medical expenditure (1 if their total out-of-pocket medical expenses are larger than 30% of their household income)
- tab5_needmet_med_inp - Outcome: Binary indicator of whether the participant feels that they received all needed medical care in past 12 months

```r
# Load in the data
data <- haven::read_dta("OHIE.dta")
```

Hint: This was an experiment with imperfect compliance. Instead of creating a "participated" or "complied" variable, simply use "treatment" as the instrument and "ohp_all_ever_admin" (enrollment in Medicaid) as the main independent variable of interest.

Question A (25 points)

Estimate the intent-to-treat effects of being selected to sign up for Medicaid on each of the four outcomes (elevated blood pressure, depression, catastrophic medical expenditure, and whether respondents had their health care needs met). Provide 95% confidence intervals for each estimate and interpret your results. (Use lm_robust)

```r
# Estimate the ITT on elevated blood pressure
lm_robust(tab2bp_hyper ~ treatment, data = data)
```

```
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.1591    0.00480  33.182 3.21e-231   0.1497   0.1685 12186
## treatment    -0.0016    0.00662  -0.242  8.09e-01  -0.0146   0.0114 12186
```

The ITT effect on elevated blood pressure is -0.0016. This indicates a tiny decrease in the outcome variable associated with the treatment. The 95% confidence interval for this estimate is [-0.0146, 0.0114], which includes 0. As a result, we would not reject the null hypothesis that being selected has no effect on blood pressure at $\alpha = .05$. This suggests that there is no significant evidence to conclude that being selected has an effect on blood pressure.

```r
# Estimate the ITT on depression
lm_robust(tab2phqtot_high ~ treatment, data = data)
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.3037    0.00603   50.33 0.00e+00    0.292   0.3155 12159
## treatment    -0.0349    0.00821   -4.26 2.09e-05   -0.051  -0.0188 12159
```

The ITT effect on depression is -0.0349. Like blood pressure, this also indicates a tiny (but larger) decrease in the outcome variable associated with the treatment. The 95% confidence interval for this estimate is [-0.051, -0.0188], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on depression at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in relieving depression.

```r
# Estimate the ITT on catastrophic expenditures
lm_robust(tab4_catastrophic_exp_inp ~ treatment, data = data)
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.0538    0.00300   17.92 6.77e-71   0.0479  0.05971 11793
## treatment    -0.0153    0.00388   -3.94 8.34e-05  -0.0229 -0.00766 11793
```

The ITT effect on catastrophic medical expenditure is -0.0153. Like blood pressure and depression, this also indicates a tiny (between them) decrease in the outcome variable associated with the treatment. The 95% confidence interval for this estimate is [-0.0229, -0.00766], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on catastrophic medical expenditure at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in reducing catastrophic medical expenditure.

```r
# Estimate the ITT on "needs met"
lm_robust(tab5_needmet_med_inp ~ treatment, data = data)
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.6124    0.00638   96.02 0.00e+00   0.5999   0.6249 12214
## treatment     0.0345    0.00875    3.94 8.19e-05   0.0173   0.0516 12214
```

The ITT effect on whether needs are met is 0.0345. Unlike the above, this indicates an increase (with a similar level to depression) in the outcome variable associated with the treatment. The 95% confidence interval for this estimate is [0.0173, 0.0516], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on whether needs are met at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in satisfying needs.

Question B (25 points)

Suppose that researchers actually wanted to estimate the effect of Medicaid enrollment (ohp_all_ever_admin) on each of the four outcomes. Suppose they first used a naive regression of each of the the outcomes on the indicator of Medicaid enrollment. Report a 95% confidence interval for each of your estimates and interpret your results. Why might these be biased estimates for the causal effect of Medicaid enrollment?

```r
# Estimate the Naive OLS effect on elevated blood pressure
lm_robust(tab2bp_hyper ~ ohp_all_ever_admin, data = data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.1634    0.00397   41.21   0.0000   0.1557  0.17122 12186
## ohp_all_ever_admin  -0.0181    0.00716   -2.52   0.0117  -0.0321 -0.00401 12186
```

The 95% confidence interval for this estimate is [-0.0321, -0.00401], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on blood pressure at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in reducing blood pressure.

```
# Estimate the Naive OLS effect on depression
lm_robust(tab2phqtot_high ~ ohp_all_ever_admin, data = data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.2713    0.00477   56.83 0.00e+00   0.2619   0.2806 12159
## ohp_all_ever_admin   0.0493    0.00924    5.34 9.52e-08   0.0312   0.0674 12159
```

The 95% confidence interval for this estimate is [0.0312, 0.0674], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on depression at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in worsening depression.

```
# Estimate the Naive OLS effect on catastrophic expenditures
lm_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin, data = data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.0489    0.00235   20.81 1.66e-94   0.0443  0.05355 11793
## ohp_all_ever_admin  -0.0107    0.00405   -2.65 8.13e-03  -0.0187 -0.00278 11793
```

The 95% confidence interval for this estimate is [-0.0187, -0.00278], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on catastrophic expenditures at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in reducing catastrophic medical expenditure.

```
# Naive OLS estimate on needs met
lm_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin, data = data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.6128    0.00522  117.40 0.00e+00   0.6026   0.6230 12214
## ohp_all_ever_admin   0.0613    0.00948    6.46 1.08e-10   0.0427   0.0799 12214
```

The 95% confidence interval for this estimate is [0.0427, 0.0799], which does not include 0. As a result, we reject the null hypothesis that being selected has no effect on whether needs are met at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in satisfying needs.

These estimates might be biased because the selection is not completely random. There is some selection bias. Other factors that can affect all these outcome variables might be unrecognized. For example, the low-income, uninsured individuals who are selected might already be more susceptible to high blood pressure, worse depression, higher catastrophic medical expenditure and their needs are less likely to be met given their initial situation.

Question C (25 points)

Suppose we were to use assignment to treatment as an instrument for actually receiving Medicaid coverage.

Consider that not everyone who was selected to apply for Medicaid actually ended up applying and receiving coverage. Likewise, some applicants who were not selected to receive the treatment nevertheless were eventually covered. What were the compliance rates (the level of Medicaid enrollment) for subjects who were selected and subjects who were not selected? Use a "first stage" regression to estimate the effect of being selected on Medicaid enrollment to estimate the compliance rates. Is the instrument of assignment-to-treatment a strong instrument for actual Medicaid enrollment?

```
# First Stage OLS
first_stage_ols <- lm_robust(ohp_all_ever_admin ~ treatment, data=data)
first_stage_ols
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)    0.145    0.00347    42.0        0    0.139    0.152 20743
## treatment      0.236    0.00589    40.1        0    0.225    0.248 20743
```

```
# null model (compliance given an intercept only model)
null_mod<-lm_robust(ohp_all_ever_admin ~ 1, data=data)

# F - Stat for Instrument Strength (use waldtest)
waldtest(first_stage_ols, null_mod,test="F")
```

```
## Wald test
##
## Model 1: ohp_all_ever_admin ~ treatment
## Model 2: ohp_all_ever_admin ~ 1
##   Res.Df Df    F Pr(>F)
## 1  20743
## 2  20744 -1 1610 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The compliance rate for subjects who were not selected is 0.145 (14.5%). The compliance for subjects who were selected is 0.145+0.236 = 0.381 (38.1%). The F-statistic is 1610, which is high. The p-value is less than 2e-16, which is very small. Both support and suggest that the instrument of assignment-to-treatment appears to be a strong instrument for actual Medicaid enrollment.

Question D (25 points)

Now estimate the effect of Medicaid enrollment on each of the four outcomes using an instrumental variables strategy. Report a 95% confidence interval for your estimates and interpret your results. Compare the estimates to those you obtained in Question B.

```
# Estimate the IV effect on elevated blood pressure (use iv_robust())
iv_robust(tab2bp_hyper ~ ohp_all_ever_admin|treatment, data=data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.1601    0.00818  19.567 5.71e-84   0.1440   0.1761 12186
## ohp_all_ever_admin  -0.0063    0.02606  -0.242 8.09e-01  -0.0574   0.0448 12186
```

The effect of Medicaid enrollment on blood pressure using an instrumental variables strategy is -0.0063. The 95% confidence interval is $[-0.0574, 0.0448]$, which includes 0. As a result, we would not reject the null hypothesis of being selected has no effect on blood pressure at $\alpha = .05$. This suggests that there is no significant evidence to conclude that being selected has an effect on blood pressure.

This estimate is different compared to the one from Question B. In Question B, the naive regression concludes that being selected reduces blood pressure. We can see clearly the naive regression is biased.

```
# Estimate the IV effect on depression
iv_robust(tab2phqtot_high ~ ohp_all_ever_admin|treatment, data=data)
```

```
##                    Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper
## (Intercept)           0.325     0.0104   31.26 2.38e-206    0.304   0.3452
## ohp_all_ever_admin   -0.138     0.0329   -4.19  2.84e-05   -0.202  -0.0732
##                       DF
## (Intercept)        12159
## ohp_all_ever_admin 12159
```

The effect of Medicaid enrollment on depression using an instrumental variables strategy is -0.138. The 95% confidence interval is [-0.202, -0.0732], which does not include 0. As a result, we reject the null hypothesis of being selected has no effect on depression at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in relieving depression.

This estimate is different compared to the one from Question B. In Question B, the naive regression concludes that being selected worsens depression. We can see clearly the naive regression is biased.

```
# Estimate the IV effect on catastrophic expenditures
iv_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin|treatment, data=data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)          0.0631    0.00508   12.43 3.03e-35   0.0532   0.0731 11793
## ohp_all_ever_admin  -0.0604    0.01543   -3.91 9.16e-05  -0.0906  -0.0301 11793
```

The effect of Medicaid enrollment on catastrophic medical expenditure using an instrumental variables strategy is -0.0604. The 95% confidence interval is [-0.0906, -0.0301], which does not include 0. As a result, we reject the null hypothesis of being selected has no effect on catastrophic medical expenditure at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in reducing catastrophic medical expenditure.

This estimate is similar compared to the one from Question B. They both conclude that being selected reduces catastrophic medical expenditure. The difference is that this estimate of IV effect is more negative, which indicates a stronger effect in reducing catastrophic medical expenditure.

```
# IV estimate on needs met
iv_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin|treatment, data=data)
```

```
##                    Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)           0.592     0.0109   54.45 0.00e+00    0.570    0.613 12214
## ohp_all_ever_admin    0.135     0.0344    3.94 8.35e-05    0.068    0.203 12214
```

The effect of Medicaid enrollment on whether needs are met using an instrumental variables strategy is 0.135. The 95% confidence interval is [0.068, 0.203], which does not include 0. As a result, we reject the null hypothesis of being selected has no effect on whether needs are met at $\alpha = .05$. This suggests that there is convincing evidence to conclude that being selected has a statistically distinguishable effect in satisfying needs.

This estimate is similar compared to the one from Question B. They both conclude that being selected satisfies needs. The difference is that this estimate of IV effect is more positive, which indicates a stronger effect in satisfying needs.