# Problem Set 4

Anita Ye - yy3557 - 001

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

```
# 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)

```
# Estimate the ITT on elevated blood pressure
itt_bp <- lm_robust(tab2bp_hyper ~ treatment, data = data)
summary(itt_bp)
```

```
##
## Call:
## lm_robust(formula = tab2bp_hyper ~ treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             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
##
## Multiple R-squared:  4.8e-06 ,   Adjusted R-squared:  -7.73e-05
## F-statistic: 0.0584 on 1 and 12186 DF,  p-value: 0.809
```

The estimated treatment effect (ITT) on elevated blood pressure is -0.0016, with a 95% confidence interval ranging from -0.0146 to 0.0114. Since this interval includes the value 0, it suggests that there is no statistically

significant impact of winning the Medicaid lottery on blood pressure levels. Consequently, we cannot reject the null hypothesis, which posits that winning the Medicaid lottery does not affect blood pressure.

```
# Estimate the ITT on depression
itt_depression <- lm_robust(tab2phqtot_high ~ treatment, data = data)
summary(itt_depression)
```

```
##
## Call:
## lm_robust(formula = tab2phqtot_high ~ treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             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
##
## Multiple R-squared:  0.00149 ,   Adjusted R-squared:  0.00141
## F-statistic: 18.1 on 1 and 12159 DF,  p-value: 2.09e-05
```

The estimated treatment effect (ITT) on depression is -0.0349, with a 95% confidence interval ranging from -0.051 to -0.0188. This interval does not include the value 0, indicating a statistically significant impact. Therefore, we can reject the null hypothesis and infer that winning the Medicaid lottery is associated with a reduction in depression.

```
# Estimate the ITT on catastrophic expenditures
itt_catastrophic_exp <- lm_robust(tab4_catastrophic_exp_inp ~ treatment, data = data)
summary(itt_catastrophic_exp)
```

```
##
## Call:
## lm_robust(formula = tab4_catastrophic_exp_inp ~ treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             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
##
## Multiple R-squared:  0.00133 ,   Adjusted R-squared:  0.00124
## F-statistic: 15.5 on 1 and 11793 DF,  p-value: 8.34e-05
```

The estimated treatment effect (ITT) on catastrophic expenditures is -0.0153, with a 95% confidence interval ranging from -0.0229 to -0.00766. This interval excludes 0, indicating a significant impact. Consequently, we can reject the null hypothesis, suggesting that winning the Medicaid lottery decreases the likelihood of individuals facing catastrophic medical expenditures.

```
# Estimate the ITT on "needs met"
itt_needs_met <- lm_robust(tab5_needmet_med_inp ~ treatment, data = data)
summary(itt_needs_met)
```

```
##
## Call:
## lm_robust(formula = tab5_needmet_med_inp ~ treatment, data = data)
##
```

```
## Standard error type:  HC2
##
## Coefficients:
##             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
##
## Multiple R-squared:  0.00127 ,   Adjusted R-squared:  0.00119
## F-statistic: 15.5 on 1 and 12214 DF,  p-value: 8.19e-05
```

Regarding the ITT on the measure of "needs met" (the perception of having received all necessary medical care in the past 12 months), it is 0.0345, with a 95% confidence interval between 0.0173 and 0.0516. This interval also does not include 0. Therefore, we can reject the null hypothesis and infer that winning the Medicaid lottery increases the probability that individuals feel they have received all required medical care.

## 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
naive_bp <- lm_robust(tab2bp_hyper ~ ohp_all_ever_admin, data = data)
summary(naive_bp)
```

```
##
## Call:
## lm_robust(formula = tab2bp_hyper ~ ohp_all_ever_admin, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                    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
##
## Multiple R-squared:  0.0005 ,    Adjusted R-squared:  0.000418
## F-statistic: 6.35 on 1 and 12186 DF,  p-value: 0.0117
```

The Naive Ordinary Least Squares (OLS) estimate for elevated blood pressure is -0.0181, with a 95% confidence interval of [-0.0321, -0.00401], which excludes 0. Therefore, we can reject the null hypothesis, suggesting that winning the Medicaid lottery has an effect on blood pressure.

```r
# Estimate the Naive OLS effect on depression
naive_depression <- lm_robust(tab2phqtot_high ~ ohp_all_ever_admin, data = data)
summary(naive_depression)
```

```
##
## Call:
## lm_robust(formula = tab2phqtot_high ~ ohp_all_ever_admin, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                    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
##
## Multiple R-squared:  0.00244 ,   Adjusted R-squared:  0.00236
## F-statistic: 28.5 on 1 and 12159 DF,  p-value: 9.52e-08
```

For depression, the Naive OLS estimate is 0.0493, with a 95% confidence interval of [0.0312, 0.0674], also excluding 0. This leads us to reject the null hypothesis and infer that Medicaid lottery winners are more likely to be diagnosed with depression.

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

```
##
## Call:
## lm_robust(formula = tab4_catastrophic_exp_inp ~ ohp_all_ever_admin,
##     data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                     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
##
## Multiple R-squared:  0.000537 ,  Adjusted R-squared:  0.000452
## F-statistic: 7.01 on 1 and 11793 DF,  p-value: 0.00813
```

In the case of catastrophic expenditures, the Naive OLS estimate is -0.0107, with a 95% confidence interval of [-0.0187, -0.00278], which does not include 0. Thus, we can conclude that winning the Medicaid lottery decreases the likelihood of participants experiencing total out-of-pocket medical expenses exceeding 30% of their household income.

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

```
##
## Call:
## lm_robust(formula = tab5_needmet_med_inp ~ ohp_all_ever_admin,
##     data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                     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
##
## Multiple R-squared:  0.0033 ,    Adjusted R-squared:  0.00322
## F-statistic: 41.7 on 1 and 12214 DF,  p-value: 1.08e-10
```

Regarding the "needs met" metric (the participants' perception of having received all necessary medical care in the past 12 months), the Naive OLS estimate is 0.0613, with a 95% confidence interval of [0.0427, 0.0799], again excluding 0. We can reject the null hypothesis and conclude that lottery winners are more likely to feel that they received all needed medical care.

However, the results of this naive regression could be biased due to potential unobserved confounders or

covariates. For example, the initial health status or socio-economic status (SES) of participants selected into the Oregon Health Plan Program, regardless of lottery outcome, could suggest other routes to program access. Naive regression assumes uncorrelated features, ignoring potential covariates and confounders. Therefore, it may not be suitable in this context. Alternative methods, such as using instrumental variables, might be necessary to address these unobserved confounders.

## 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 <- lm_robust(ohp_all_ever_admin ~ treatment, data = data)
summary(first_stage)
```

```
##
## Call:
## lm_robust(formula = ohp_all_ever_admin ~ treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             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
##
## Multiple R-squared:  0.0719 ,    Adjusted R-squared:  0.0718
## F-statistic: 1.61e+03 on 1 and 20743 DF,  p-value: <2e-16
```

Based on the OLS model, the compliance rate of the selected is 0.236, which is the coefficient of the treatment. It means that being selected in the lottery increases the probability of enrolling in Medicaid by 23.6 percentage points. The compliance rate of 23.6% can be interpreted as the effectiveness of the lottery in increasing Medicaid enrollment among those selected. This is a substantial increase, indicating that the lottery selection had a significant impact on Medicaid enrollment among participants. It refers to how the chance of enrolling in Medicaid changes due to the treatment (lottery selection).

The intercept represents the estimated probability of enrolling in Medicaid for individuals who were not selected in the lottery (treatment = 0). It indicates that, in the absence of being selected for the lottery, there is a 14.5% probability of enrolling in Medicaid.

```
# Compliance rate for treated group
compliance_treated <- mean(data$ohp_all_ever_admin[data$treatment == 1])
# Compliance rate for untreated group
compliance_untreated <- mean(data$ohp_all_ever_admin[data$treatment == 0])

# 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, null_mod)
```

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

The p-value for the treatment effect is less than 0.05 (indeed, it's 0), which indicates that the effect of being selected in the lottery on Medicaid enrollment is statistically significant. The 95% confidence interval for the treatment effect ranges from 22.5% to 24.8%. This means we can be 95% confident that the true effect of being selected in the lottery on Medicaid enrollment lies within this range.

The analysis reveals that among individuals who did not win the lottery, 14.5% were still able to enroll in the Oregon Health Plan (OHP). In contrast, the enrollment rate for lottery winners increased to 38.1%. This suggests that winning the lottery enhances the likelihood of enrollment by 23.6%. It implies that approximately 23.6% of participants decided to enroll in Medicaid as a direct consequence of winning the lottery.

The large t-value (40.1) and the significant F-statistic (from the waldtest) suggest that the model is a good fit and the instrument (lottery selection) is strong and reliably predicts Medicaid enrollment.

Furthermore, the new F-statistic for the first stage of the analysis is 1610, significantly surpassing the usual thresholds for instrument strength. This robustness can be attributed to the large sample size and the substantial effect of the lottery on Medicaid enrollment, making it a strong instrumental variable 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.

```r
# Estimate the IV effect on elevated blood pressure (use iv_robust())
iv_bp <- iv_robust(
  formula = tab2bp_hyper ~ ohp_all_ever_admin | treatment,
  data = data
)
summary(iv_bp)
```

```
##
## Call:
## iv_robust(formula = tab2bp_hyper ~ ohp_all_ever_admin | treatment,
##     data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                    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
##
## Multiple R-squared:  0.000288 ,  Adjusted R-squared:  0.000206
## F-statistic: 0.0584 on 1 and 12186 DF,  p-value: 0.809
```

The Instrumental Variable (IV) estimate for the effect of Medicaid enrollment on blood pressure is -0.0063, with a 95% confidence interval of [-0.0574, 0.0448], which includes 0. Consequently, we cannot reject the null hypothesis that Medicaid enrollment does not significantly affect blood pressure. This finding contrasts with the earlier Naive OLS estimate, which suggested a reduction in blood pressure due to Medicaid enrollment. The discrepancy suggests that the Naive OLS results might be influenced by unobserved confounders, leading to the conclusion that Medicaid enrollment does not have a discernible impact on blood pressure.

```
# Estimate the IV effect on depression
iv_depression <- iv_robust(
  formula = tab2phqtot_high ~ ohp_all_ever_admin | treatment,
  data = data
)
summary(iv_depression)
```

```
##
## Call:
## iv_robust(formula = tab2phqtot_high ~ ohp_all_ever_admin | treatment,
##     data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                     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
##
## Multiple R-squared:  -0.0326 ,   Adjusted R-squared:  -0.0327
## F-statistic: 17.5 on 1 and 12159 DF,  p-value: 2.84e-05
```

For depression, the IV estimate is -0.138, with a 95% confidence interval of [-0.202, -0.0732], excluding 0. Therefore, we reject the null hypothesis and infer that Medicaid enrollment decreases diagnosed depression by 13.8%. This conclusion is the opposite of the Naive OLS estimate, which indicated a positive effect of Medicaid enrollment on depression. The IV analysis, however, suggests a negative impact, highlighting the importance of addressing potential confounders in such estimations.

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

```
##
## Call:
## iv_robust(formula = tab4_catastrophic_exp_inp ~ ohp_all_ever_admin |
##     treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                     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
```

```
##
## Multiple R-squared:  -0.011 ,    Adjusted R-squared:  -0.011
## F-statistic: 15.3 on 1 and 11793 DF,  p-value: 9.16e-05
```

Regarding catastrophic expenditures, the IV estimate is -0.0604, with a 95% confidence interval of [-0.0906, -0.0301], also not including 0. We thus reject the null hypothesis and conclude that Medicaid enrollment reduces catastrophic expenditures by 6.04%. Both the IV and Naive OLS estimates point in the same direction and have confidence intervals that exclude 0, with the IV estimate being slightly larger.

```r
# IV estimate on needs met
iv_needs_met <- iv_robust(
  formula = tab5_needmet_med_inp ~ ohp_all_ever_admin | treatment,
  data = data
)
summary(iv_needs_met)
```

```
##
## Call:
## iv_robust(formula = tab5_needmet_med_inp ~ ohp_all_ever_admin |
##     treatment, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##                    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
##
## Multiple R-squared:  -0.00154 ,  Adjusted R-squared:  -0.00162
## F-statistic: 15.5 on 1 and 12214 DF,  p-value: 8.35e-05
```

For the metric of whether participants feel they received all needed medical care, the IV estimate is 0.135, with a 95% confidence interval of [0.068, 0.203], excluding 0. This leads to the rejection of the null hypothesis, indicating that Medicaid enrollment increases participants' perception of meeting their medical needs by 13.5%. Both the IV and Naive OLS estimates align in direction and have confidence intervals that exclude 0, with the IV estimate and confidence interval being slightly larger than those of the Naive OLS.