PLSC 30600: Problem Set 2

Solutions

April 16, 2022

*This problem set is due at **11:59 pm on Thursday, April 28th**.*

Please upload your solutions as a .pdf file saved as "Yourlastname_Yourfirstinitial_pset2.pdf". In addition, an electronic copy of your .Rmd file (saved as "Yourlastname_Yourfirstinitial_pset2.Rmd") must be submitted to the course website at the same time. We should be able to run your code without error messages. In addition to your solutions, please submit an annotated version of this ‘.rmd‘ file saved as "Yourlastname_Yourfirstinitial_pset2_feedback.rmd" and a corresponding PDF saved as "Yourlastname_Yourfirstinitial_pset2_feedback.pdf" noting the problems where you needed to consult the solutions and why along with any remaining questions or concerns about the material. In order to receive credit, homework submissions must be substantially started and all work must be shown. Late assignments will not be accepted. In total your submissions should consist of four files.

Problem 1

In this problem we will revisit the Hyde (2003) study of election observers in Armenia that you examined in Problem Set 1.

For the purposes of this problem, you will be using the `armenia2003.dta` dataset

The R code below will read in this data (which is stored in the STATA .dta format)

```
### Hyde (2007) Armenia dataset
armenia <- read_dta("armenia2003.dta")
```

See Problem Set 1 for a full description of the data. The relevant columns in the dataset are:

- `kocharian` - Round 1 vote share for the incumbent (Kocharian)
- `mon_voting` - Whether the polling station was monitored in round 1 of the election
- `turnout` - Proportion of registered voters who voted in Round 1
- `totalvoters` - Total number of registered voters recorded for the polling station
- `total` - Total number of votes cast in Round 1
- `urban` - Indicator for whether the polling place was in an urban area (0 = rural, 1 = urban)
- `nearNagorno` - Indicator for whether the polling place is near the Nagorno-Karabakh region (0 = no, 1 = yes)

Part A

Divide the sample into five strata based on the total number of registered voters at each polling station (`totalvoters`):

Stratum	Total Registered Voters
Tiny	`totalvoters` < 430
Small	$430 \leq$ `totalvoters` < 1192
Medium	$1192 \leq$ `totalvoters` < 1628

Stratum	Total Registered Voters
Large	$1628 \leq \texttt{totalvoters} < 1879$
Huge	$1879 \leq \texttt{totalvoters}$

Estimate the average treatment effect of election monitoring in round 1 on incumbent vote share using a stratified difference-in-means estimator, stratifying on the total number of registered voters. Provide a 95% asymptotic confidence interval and interpret your results. Can we reject the null of no average treatment effect at the $\alpha = 0.05$ level? Compare your answer to the unadjusted estimate from Problem Set 1 and discuss why they differ.

First, make the strata

```
armenia <- armenia %>% mutate(voteStrat = case_when(totalvoters < 430 ~ "Tiny",
                                                    totalvoters >= 430&totalvoters<1192 ~ "Small",
                                                    totalvoters >= 1192&totalvoters<1628 ~ "Medium",
                                                    totalvoters >= 1628&totalvoters < 1879 ~ "Large",
                                                    totalvoters >= 1879 ~ "Huge"))
```

First, remember the unadjusted estimate

```
lm_robust(kocharian ~ mon_voting, data=armenia)
```

```
##                 Estimate  Std. Error  t value     Pr(>|t|)    CI Lower
## (Intercept)  0.54190315 0.006671177 81.23052 0.000000e+00  0.52881890
## mon_voting  -0.05867601 0.009793929 -5.99106 2.521806e-09 -0.07788495
##                 CI Upper   DF
## (Intercept)  0.55498740 1762
## mon_voting  -0.03946707 1762
```

Estimate the ATE conditional on the strata - estimate each conditional ATE and aggregate up w.r.t. the distribution of the strata.

```
lm_lin(kocharian ~ mon_voting, covariates = ~voteStrat, data=armenia)
```

```
##                                  Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)                   0.525826976 0.006141297 85.6214863 0.000000e+00
## mon_voting                   -0.016991582 0.010028364 -1.6943523 9.037588e-02
## voteStratLarge_c              0.024265835 0.018821196  1.2892823 1.974700e-01
## voteStratMedium_c             0.004106563 0.019961811  0.2057210 8.370327e-01
## voteStratSmall_c              0.060937516 0.018045104  3.3769556 7.489430e-04
## voteStratTiny_c               0.179722268 0.018646630  9.6383245 1.854359e-21
## mon_voting:voteStratLarge_c  -0.016974436 0.025808671 -0.6577028 5.108154e-01
## mon_voting:voteStratMedium_c  0.010504668 0.027239784  0.3856370 6.998123e-01
## mon_voting:voteStratSmall_c   0.003573102 0.030293090  0.1179511 9.061199e-01
## mon_voting:voteStratTiny_c    0.029886008 0.033750324  0.8855028 3.760069e-01
##                                  CI Lower    CI Upper   DF
## (Intercept)                   0.51378194 0.537872009 1754
## mon_voting                   -0.03666039 0.002677223 1754
## voteStratLarge_c             -0.01264850 0.061180175 1754
## voteStratMedium_c            -0.03504488 0.043258010 1754
## voteStratSmall_c              0.02554534 0.096329692 1754
## voteStratTiny_c               0.14315031 0.216294227 1754
## mon_voting:voteStratLarge_c  -0.06759343 0.033644560 1754
## mon_voting:voteStratMedium_c -0.04292120 0.063930530 1754
## mon_voting:voteStratSmall_c  -0.05584126 0.062987466 1754
## mon_voting:voteStratTiny_c   -0.03630909 0.096081106 1754
```

After adjusting for the size of the polling station, we estimate that monitoring reduced incumbent vote share by 1.7pp - a significant change from the 5.9pp effect that we obtain without adjustment. Furthermore, the 95% confidence interval $(-0.036, 0.0027)$ contains zero which means that we would fail to reject the null of no treatment effect at the $\alpha = .05$ level. Adjusting for the size of the polling location (total number of registered voters) explains away a significant chunk of the observed difference between monitored and non-monitored locations. This is likely because the treatment was actually assigned non-randomly and smaller locations (which may have been harder to reach) were less likely to receive monitors. Likewise, these locations

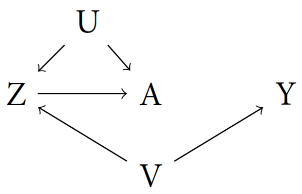
Part B

In Table 4 of the paper, Hyde uses an estimator for the average treatment effect of a polling place receiving election monitors in round 1 on the incumbent's vote share in round 1 *conditional* on the total number of votes cast in the election (`total`). Will this approach be unbiased for the average treatment effect of election monitors on the incumbent's vote share if we believe that one of the mechanisms through which election monitoring operates is by reducing the incidence of ballot-stuffing (which inflates the number of "cast" votes in the election)? Why or why not?

If election monitoring affects the total number of votes cast, then this variable is post-treatment. Adjusting for it risks inducing "collider" bias if there are other factors associated with number of votes cast and incumbent vote share (e.g. incumbent performs better in districts with a smaller number of cast votes).

Problem 2

Consider the following causal directed acyclic graph:



Part A

List all of the paths from A to Y and identify those paths as causal or noncausal.

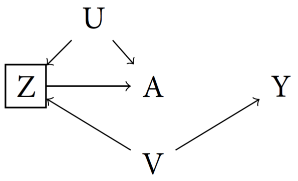
---

1. $Y \leftarrow V \rightarrow Z \rightarrow A$ - Non-causal
2. $Y \leftarrow V \rightarrow Z \leftarrow U \rightarrow A$ - Non-causal

Part B

Given the DAG in Figure 1, are A and Y dependent?

Yes, although path 2 is blocked by the collider at Z, path 1 is unblocked absent conditioning.

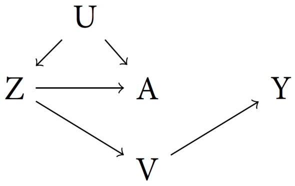


Part C

Suppose that we control for Z (by regression, subclassification, etc), are A and Y dependent given Z?

Yes, conditioning on Z blocks path 1 but opens up path 2 as Z is a collider on this path.

Part D



Suppose now that we flip the direction of the arrow from V to Z, so that Z $\implies$ V (Figure 3). In this revised DAG, are A and Y dependent?

Yes, both of the non-causal paths are unblocked (no colliders).

Part E

Suppose in this revised DAG we now control for Z. Are A and Y dependent given Z?

No, conditioning on Z blocks both of the paths from A to Y and Z is no longer a collider on the second path. Therefore Z d-separates A and Y.

Problem 3

In new democracies and post-conflict settings, Truth and Reconciliation Commissions (TRCs) are often tasked with investigating and reporting about wrongdoing in previous governments. Depending on the context, institutions such as TRCs are expected to reduce conflict (e.g. racial hostilities) and promote peace.

In 1995, South Africa's new government formed a national TRC in the aftermath of apartheid. Gibson 2004 uses survey data collected from 2000-2001 to examine whether this TRC promoted inter-racial reconciliation. The outcome of interest is respondent racial attitudes (as measured by the level of agreement with the prompt: "I find it difficult to understand the customs and ways of [the opposite racial group]".) The treatment is "exposure to the TRC" as measured by the individual's level of self-reported knowledge about the TRC.

You will need to use the `trc_data.dta` file for this question. The code below reads the data into R

```
trc <- read_dta("trc_data.dta")
```

The relevant variables are:

- `RUSTAND` - Outcome: respondent's racial attitudes (higher values indicate greater agreement)

- `TRCKNOW` - Treatment dummy (1 = if knows about the TRC, 0 = otherwise)
- `age` - Respondent age (in 2001)
- `female` - Respondent gender
- `wealth` - Measure of wealth constructed based on asset ownership (assets are fridge, floor polisher, vacuum cleaner, microwave oven, hi-fi, washing machine, telephone, TV, car)
- `religiosity` - Self-reported religiosity (7 point scale)
- `ethsalience` - Self-reported ethnic identification (4 point scale)
- `rcblack` - Respondent is black
- `rcwhite` - Respondent is white
- `rccol` - Respondent is coloured (distinct multiracial ethnic group)
- `EDUC` - Level of education (9 point scale)

Part A

Estimate the average treatment effect of TRC exposure on respondents' racial attitudes under the assumption that TRC exposure is completely ignorable. Report a 95% confidence interval for your estimate and interpret your results.

```r
library(estimatr)
# Load in the TRC data (it's a STATA .dta so we use the haven package)
TRC_data <- haven::read_dta("trc_data.dta")

# Baseline results
naive_reg <- lm_robust(RUSTAND ~ TRCKNOW, data=TRC_data)

# What's the point estimate/SE
summary(naive_reg)
```

```
##
## Call:
## lm_robust(formula = RUSTAND ~ TRCKNOW, data = TRC_data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)   2.5311    0.02806  90.212 0.000e+00   2.4761   2.5862 3203
## TRCKNOW      -0.2177    0.04433  -4.911 9.492e-07  -0.3047  -0.1308 3203
##
## Multiple R-squared:  0.007626 ,   Adjusted R-squared:  0.007316
## F-statistic: 24.12 on 1 and 3203 DF,  p-value: 9.492e-07
```

We estimate that knowing about the TRC is associated with a -.2177 point decrease, on average, in respondents' hostile racial attitudes (as measured by agreement with the phrase "I find it difficult to understand the customs and ways of [the opposite racial group]"). The 95% confidence interval is [-.3047, -.1308] and we would reject the null of no effect at the .05 level.

Part B

Examine whether exposed and nonexposed respondents differ on the full set of observed covariates using a series of balance tests. In which ways do exposed and nonexposed respondents differ? What does this tell you about whether the assumption that TRC exposure is completely ignorable is reasonable?

```
# Let's check for balance between treatment and control on our
TRC_data %>% group_by(TRCKNOW) %>% summarize(age = mean(age), female = mean(female), wealth = mean(weal
```

```
## # A tibble: 2 x 10
##   TRCKNOW   age female wealth  educ religiosity ethsalience rcblack rcwhite
##     <dbl> <dbl>  <dbl>  <dbl> <dbl>       <dbl>       <dbl>   <dbl>   <dbl>
## 1       0  40.5  0.433  5793.  3.85        3.92        2.71   0.513   0.253
## 2       1  38.9  0.538  6945.  4.29        3.84        2.73   0.552   0.270
## # ... with 1 more variable: rcmult <dbl>
```

On average, exposed respondents tend to be slightly younger (by about 2 years), more likely to be male, slightly wealthier and more educated. Exposed respondents are more likely to be black, but slightly less likely to be of mixed racial ancestry.

Interestingly, exposed and unexposed respondents do not seem to significantly differ on self-reported ethnic identification (in other words, how salient ethnicity is to them), which might be a strong predictor of the outcome of interest.

Overall, the magnitude of the discrepancy between treated and control on these observed covariates strongly suggests that a strict ignorability assumption is not reasonable. Exposure to the Truth and Reconciliation Commission is not randomly assigned and exposed respondents likely differ significantly from those who were not exposed in ways that would likely affect the outcome of interest.

Part C

Now assume that TRC exposure is conditionally ignorable given the set of observed covariates. Use an additive logistic regression model to estimate the propensity score for each observation. With this model, construct inverse propensity of treatment weights (IPTW) for each observation and compute a point estimate for the ATE.

```
#model for outcome variable (just regress outcome on treatment)
outcome_model <- RUSTAND ~ TRCKNOW
#model for treatment variable
treatment_model <- TRCKNOW ~ age + female + wealth + religiosity +
  ethsalience + rcblack + rcwhite + rccol + EDUC
#compute propensity scores
TRC_data$pscores <- glm(treatment_model, family = binomial(), data = TRC_data)$fitted.values

# Compute the weights
TRC_data$w <- TRC_data$TRCKNOW*(mean(TRC_data$TRCKNOW)/TRC_data$pscores) +
  (1-TRC_data$TRCKNOW)*(mean(1-TRC_data$TRCKNOW)/(1-TRC_data$pscores))
# Fit the outcome model
fit.ipsw <- lm_robust(outcome_model, weights=TRC_data$w, data=TRC_data)
ipw_coef <- summary(fit.ipsw)$coefficients[2,1]
ipw_coef
```

```
## [1] -0.1631028
```

Adjusting via IP weighting, we find that knowing about the TRC is associated with a -.163 point decrease, on average, in respondents' hostile racial attitudes.

Part D

Using a pairs bootstrap (resampling individual rows of the data with replacement), obtain estimate for the standard error of your IPTW estimator for the ATE. Compute a 95% confidence interval and interpret your findings. Compare your results in Parts C/D to your estimate from Part A and discuss.

```
# Set random seed
set.seed(10003)

#IPTW Bootstrap
n_iter <- 1000 # 1000 iterations
iptw_boot <- rep(NA, n_iter)

# For each iteration
for (j in 1:n_iter){
  # Resample the data (with replacement)
  TRC_data_boot <- TRC_data[sample(nrow(TRC_data), replace=T),]
  # Fit the model on the resampled data
  TRC_data_boot$pscores_boot <- glm(treatment_model, family = binomial(),
                                    data = TRC_data_boot)$fitted.values
  # Calculate the weights
  TRC_data_boot$w <- TRC_data_boot$TRCKNOW*(mean(TRC_data_boot$TRCKNOW)/TRC_data_boot$pscores_boot) +
  (1-TRC_data_boot$TRCKNOW)*(mean(1 - TRC_data_boot$TRCKNOW)/(1-TRC_data_boot$pscores_boot))
  # Take the difference-in-means
  iptw_boot[j] <- summary(lm_robust(outcome_model, weights=TRC_data_boot$w,
                                    data=TRC_data_boot))$coefficients[2,1]
}
# Take the SD of the bootstrapped sampling distribution to estimate our SE
iptw_se <- sd(iptw_boot)
iptw_se
```

```
## [1] 0.04451084
```

```
# Compute a 95% confidence interval
ipw_95_CI <- c(ipw_coef - qnorm(.975)*iptw_se, ipw_coef + qnorm(.975)*iptw_se)
ipw_95_CI
```

```
## [1] -0.25034242 -0.07586313
```

The bootstrapped standard error is 0.0445, yielding a 95% confidence interval of [-0.2503, -0.0759]. We would still reject the null of no effect at the .05 level even after adjusting for the covariates. However, our point estimate is somewhat more attenuated towards zero, suggesting that some of the original difference we observed in Part A between the two groups is attributable to confounding driven by these variables.

Part E

Now, instead of weighting, we will consider stratification on the propensity score directly.

Based on its estimated propensity score, assign each observation to one of six equally-sized strata (bins). Examine the stratum with the highest propensity scores and, within that stratum, carry out a series of balance tests between exposed and nonexposed respondents for the full set of observed covariates. How does the balance within this particular stratum compare to the overall balance you found in part B?

```
# Take propensity scores from before, find cutpoints
pscore_cutpoints <- quantile(TRC_data$pscores, seq(0, 1, by=1/6))
# Lowest bin bounded by 0, highest bounded by 100 (this avoids weird issues where
# the extreme observations don't get binned)
pscore_cutpoints[1] <- 0
pscore_cutpoints[7] <- 1
# Use "cut" to assign each observation to a bin
TRC_data$pscoreStrat <- cut(TRC_data$pscores, pscore_cutpoints, labels= F)
## Validate the split is roughly even
table(TRC_data$pscoreStrat)
```

```
##
##   1   2   3   4   5   6
## 535 534 534 534 534 534
```

```
## Look at the highest stratum - run a balance check
TRC_data %>% filter(pscoreStrat == 6) %>% group_by(TRCKNOW) %>% summarize(age = mean(age), female = mea
```

```
## # A tibble: 2 x 10
##   TRCKNOW   age female  educ wealth religiosity ethsalience rcblack rcwhite
##     <dbl> <dbl>  <dbl> <dbl>  <dbl>       <dbl>       <dbl>   <dbl>   <dbl>
## 1       0  38.9  0.725  5.70 12399.        4.09        2.73   0.392   0.487
## 2       1  39.1  0.728  5.75 11926.        3.87        2.76   0.443   0.441
## # ... with 1 more variable: rcmult <dbl>
```

We find that within that propensity score stratum, the magnitude of the differences in covariates between treated and control is substantially reduced. For example, the discrepancy between the mean age of respondents in treatment and respondents in control has been reduced to less than 1. There is also essentially no difference in mean education within this stratum. There may be some residual imbalance, but just by inspection of the treated/control means, the imbalance within the stratum is substantially smaller than imbalance in the sample overall.

Part F

Estimate the average treatment effect using a stratified difference-in-means estimator based on your strata from Part E. Use the typical stratified variance estimator (don't bootstrap here) and report a 95% confidence interval. Compare your results to your findings in Part A and your results from D.

```
# Get the point estimate using a stratified estimator
# Stratifying on pscoreStrat varaible
# lm_lin implements the de-meaned covariate "Lin" estimator
# You could also\ do this manually!
strat_reg <- lm_lin(outcome_model,
                    covariates = ~ as.factor(pscoreStrat),
                    data = TRC_data)

strat_reg
```

```
##                                          Estimate Std. Error    t value
## (Intercept)                            2.51817939 0.02875183 87.5832689
## TRCKNOW                               -0.17602937 0.04567931 -3.8535909
## (as.factor(pscoreStrat)2)_c            0.02131783 0.08481971  0.2513311
## (as.factor(pscoreStrat)3)_c            0.26604255 0.08758315  3.0375996
## (as.factor(pscoreStrat)4)_c            0.15836102 0.08995117  1.7605220
## (as.factor(pscoreStrat)5)_c            0.28314622 0.09774372  2.8968226
## (as.factor(pscoreStrat)6)_c           -0.31733727 0.10288149 -3.0844932
## TRCKNOW:(as.factor(pscoreStrat)2)_c -0.17402053 0.16626502 -1.0466455
## TRCKNOW:(as.factor(pscoreStrat)3)_c -0.22100770 0.16212324 -1.3632080
## TRCKNOW:(as.factor(pscoreStrat)4)_c -0.25913239 0.16447785 -1.5754850
## TRCKNOW:(as.factor(pscoreStrat)5)_c -0.26233899 0.16318807 -1.6075868
## TRCKNOW:(as.factor(pscoreStrat)6)_c -0.15855384 0.16292986 -0.9731417
##                                        Pr(>|t|)    CI Lower    CI Upper   DF
## (Intercept)                         0.000000000  2.46180547  2.57455332 3193
## TRCKNOW                             0.000118688 -0.26559312 -0.08646562 3193
## (as.factor(pscoreStrat)2)_c         0.801574321 -0.14498879  0.18762445 3193
## (as.factor(pscoreStrat)3)_c         0.002403994  0.09431763  0.43776747 3193
## (as.factor(pscoreStrat)4)_c         0.078415059 -0.01800689  0.33472893 3193
## (as.factor(pscoreStrat)5)_c         0.003795267  0.09149940  0.47479304 3193
## (as.factor(pscoreStrat)6)_c         0.002056439 -0.51905776 -0.11561678 3193
## TRCKNOW:(as.factor(pscoreStrat)2)_c 0.295342328 -0.50001755  0.15197649 3193
## TRCKNOW:(as.factor(pscoreStrat)3)_c 0.172913078 -0.53888390  0.09686851 3193
## TRCKNOW:(as.factor(pscoreStrat)4)_c 0.115243596 -0.58162530  0.06336052 3193
## TRCKNOW:(as.factor(pscoreStrat)5)_c 0.108024569 -0.58230301  0.05762503 3193
## TRCKNOW:(as.factor(pscoreStrat)6)_c 0.330556595 -0.47801158  0.16090391 3193
```

Adjusting via stratification on the propensity scores, we find that knowing about the TRC is associated with a -.176 point decrease, on average, in respondents' hostile racial attitudes. The estimated standard error is 0.0457, yielding a 95% confidence interval of [-0.266, -0.0846]. We would still reject the null of no effect at the .05 level even after adjusting for the covariates using this method. The point estimate using stratification is nearly the same as the estimate using IPTW, suggesting that these two different approaches to adjusting for confounding yield very similar results (as we might expect since they rely on the same propensity scores as inputs). Again, the estimate differs from the unadjusted estimate in A by being somewhat closer to 0, suggesting that the observed confounders are driving some (but not necessarily all) of the difference between exposed and unexposed respondents.

Problem 4

Consider an experiment with N units. Each unit i in the sample belongs to one of G mutually exclusive strata. $G_i = g$ denotes that the ith unit belongs to stratum g. N_g denotes the size of stratum g and $N_{t,g}$ denotes the number of treated units in that stratum. Suppose that treatment is assigned via complete randomization within each block. Within each stratum, $N_{t,g}$ units are randomly selected to receive treatment and the remainder: $N_{c,g} = N_g - N_{t,g}$ receive control. Assume that the proportion of treated units in each stratum, $\frac{N_{t,g}}{N_g}$, varies depending on the stratum but is a known constant. After treatment is assigned, you record an outcome Y_i for each unit in the sample. Assume consistency holds with respect to the potential outcomes:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

. In total, there are $N_t = \sum_{g=1}^{G} \frac{N_{t,g}}{N_g}$ treated units and $N_c = \sum_{g=1}^{G} \frac{N_{c,g}}{N_g}$ control units.

The probability that a unit i receives treatment can be written as function of its group membership: $Pr(D_i = 1|G_i) = \frac{N_{t,G_i}}{N_{G_i}}$.

Consider finite sample inference for the SATE τ (no sampling from a super-population, potential outcomes are fixed)

$$\tau = \frac{1}{N} \sum_{i=1}^{N} Y_i(1) - Y_i(0)$$

Part A

Show that if $\frac{N_{t,g}}{N_g}$ is not the same for all g, the simple difference-in-means estimator $\hat{\tau}$ is biased for the SATE.

$$\hat{\tau} = \frac{1}{N_t}\sum_{i=1}^{N} Y_i D_i - \frac{1}{N_c}\sum_{i=1}^{N} Y_i(1-D_i)$$

Hint: Take the expectation of $\hat{\tau}$ conditional on potential outcomes $\mathbf{Y}(1)$, $\mathbf{Y}(0)$ and known group indicators $\mathbf{G}$.

Start by taking the expectation conditional on the potential outcomes and group indicators and apply linearity of expectations

$$E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] = \frac{1}{N_t}\sum_{i=1}^{N} E[Y_i D_i|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] - \frac{1}{N_c}\sum_{i=1}^{N} E[Y_i(1-D_i)|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}]$$

By consistency and using the fact that the potential outcomes are a constant

$$E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] = \frac{1}{N_t}\sum_{i=1}^{N} Y_i(1)E[D_i|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] - \frac{1}{N_c}\sum_{i=1}^{N} Y_i(0)E[(1-D_i)|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}]$$

Then, we know the probability of treatment conditional on the group membership indicator G_i

$$E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] = \frac{1}{N_t}\sum_{i=1}^{N} Y_i(1)\frac{N_{t,G_i}}{N_{G_i}} - \frac{1}{N_c}\sum_{i=1}^{N} Y_i(0)\frac{N_{c,G_i}}{N_{G_i}}$$

Combine terms (get a single sum)

$$E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] = \frac{1}{N}\sum_{i=1}^{N} Y_i(1)\frac{N}{N_t}\frac{N_{t,G_i}}{N_{G_i}} - Y_i(0)\frac{N}{N_c}\frac{N_{c,G_i}}{N_{G_i}}$$

Rearrange terms for clarity

$$E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] = \frac{1}{N}\sum_{i=1}^{N} Y_i(1)\left(\frac{N_{t,G_i}/N_{G_i}}{N_t/N}\right) - Y_i(0)\left(\frac{N_{c,G_i}/N_{G_i}}{N_c/N}\right)$$

Intuitively, this is only equal to the SATE if the proportion of treated units in each group is equal to the overall proportion of treated units. Otherwise, $\hat{\tau}$ is biased for the true SATE:

$$\text{Bias}(\hat{\tau}) = E[\hat{\tau}|\mathbf{Y}(1),\mathbf{Y}(0),\mathbf{G}] - \tau_{\text{SATE}} = \frac{1}{N}\sum_{i=1}^{N} Y_i(1)\left(\frac{N_{t,G_i}/N_{G_i}}{N_t/N} - 1\right) - Y_i(0)\left(\frac{N_{c,G_i}/N_{G_i}}{N_c/N} - 1\right)$$

Part B

Consider the following weighted estimator where $w_i^{(1)}$ and $w_i^{(0)}$ are known constant weights for each observation:

$$\hat{\tau}_w = \frac{1}{N_t}\sum_{i=1}^{N} Y_i D_i w_i^{(1)} - \frac{1}{N_c}\sum_{i=1}^{N} Y_i(1-D_i)w_i^{(0)}$$

Using your result from Part A, find an expression for both $w_i^{(1)}$ and $w_i^{(0)}$ that makes $\hat{\tau}_w$ unbiased for the SATE. Interpret the weights substantively - what do they represent?

Hint: Your weights will be a function of the group membership indicator G_i.

The weights are the inverse treatment propensities (normalized by the number of treated/control units):

$$w_i^{(1)} = \frac{N_{G_i}}{N_{t,G_i}} \times \frac{N_t}{N}$$

$$w_i^{(0)} = \frac{N_{G_i}}{N_{c,G_i}} \times \frac{N_c}{N}$$

From our proof above, we have

$$E[\hat{\tau}_w | \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{G}] = \frac{1}{N_t} \sum_{i=1}^{N} Y_i(1) w_i^{(1)} \frac{N_{t,G_i}}{N_{G_i}} - \frac{1}{N_c} \sum_{i=1}^{N} Y_i(0) w_i^{(0)} \frac{N_{c,G_i}}{N_{G_i}}$$

Substituting in the weights and cancelling

$$E[\hat{\tau}_w | \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{G}] = \frac{1}{N_t} \sum_{i=1}^{N} Y_i(1) \frac{N_t}{N} - \frac{1}{N_c} \sum_{i=1}^{N} Y_i(0) \frac{N_c}{N}$$

$$E[\hat{\tau}_w | \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{G}] = \frac{1}{N} \sum_{i=1}^{N} Y_i(1) - \frac{1}{N} \sum_{i=1}^{N} Y_i(0)$$

Which yields the SATE.

$$E[\hat{\tau}_w | \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{G}] = \frac{1}{N} \sum_{i=1}^{N} Y_i(1) - Y_i(0)$$