problem set 1 - group 4

Problem 1

```r
# create the dataset

y <- c(rep(0, 500), rep(1, 300), rep(0, 60), rep(1, 40), rep(0, 20), rep(1, 30))
d <- c(rep(0, 500), rep(0, 300), rep(0, 60), rep(0, 40), rep(1, 20), rep(1, 30))
z <- c(rep(0, 500), rep(0, 300), rep(1, 60), rep(1, 40), rep(1, 20), rep(1, 30))

df <- cbind(y, d, z)

# a - calculate the itt

mean_control <- mean(y[z == 0])

mean_treatment <- mean(y[z == 1])

itt <- mean_treatment - mean_control

print(itt) # the itt is 0.0916667
```

```
## [1] 0.09166667
```

```r
# b - 50 compliers (since there are no always takers.)

# c - among z = 0 (assigned to control), how many are compliers ?
# we can't make the calculation, we can only estimate it.
# there can be never takers and we don't see the difference.

# d - proportion of compliers
# now we need to get the denominator, which is the proportion of compliers

# first we estimate share of never-takers among those assigned to treatment
share_nev_treated <- 100/150

# now we need to estimate the number of compliers among assigned to control
compliers_incont <- 800 - (share_nev_treated*800)

# now we get the estimated share of compliers in the study
tot_compliers <- compliers_incont + 50

# d - 0.3333
print(share_complier <- tot_compliers/950)
```

```
## [1] 0.3333333
```

```r
# now we calculate the CACE estimate
ca_ce <- itt/share_complier

print(ca_ce)  # 0.275
```

```
## [1] 0.275
```

Problem 2

Iyer (2010) is interested in the effect of direct rule by the British during the colonial period on public goods provision in India, compared to indirect rule through native or "princely" states. Lord Dalhousie, Governor-General of India from 1848 to 1856, announced that:

On all occasions where heirs natural shall fail, the territory should be made to lapse and adoption should not be permitted, excepting in those cases in which some strong political reason may render it expedient to depart from this general rule.

In other words, districts in which a native ruler lacking an heir died during the period of Dalhousie's rule should be annexed by the British, according to this "Doctrine of Lapse" policy. Iyer argues that the death of an heirless ruler in the period of Dalhousie's rule can be used as an instrumental variable for direct colonial rule.

(a) What two groups of units would you compare when doing an intent-to-treat analysis here? Numerically, how is intent-to-treat analysis related to instrumental-variables analysis?

For an intent to treat analysis, we would use the variable that is purportedly "as good as random" for the treatment assignment variable, in this case the death of an heirless ruler in the period of Dalhousie's rule. The outcome variable is public goods provision. We would compare the (public good provisions of) the group of units where there was a death of a native ruler without an heir, to the group of units where there was not a death of native ruler or where there was the death of one, but there was an heir.

Here, the instrumental-variables analysis is equivalent to the CACE (in the bivariate, binary treatment case). Numerically, the intent-to-treat analysis is the numerator of this.

(b) Define Compliers, Always-Treats, Never-Treats, and Defiers in this context.

Compliers: Those units where the native ruler died without an heir, and British direct rule was taken up. Those units where the native ruler died with an heir (or where there was no death of the ruler), and British direct rule was not imposed - it remained a princely state.

Always-Treats: Those units where the native ruler died without an heir, and British direct rule was taken up. Those units where the native ruler died with an heir (or where there was no death of the ruler), and British direct rule was taken up.

Never-Treats: Those units where the native ruler died without an heir, and British direct rule was not taken up. Those units where the native ruler died with an heir (or where there was no death of the ruler), and British direct rule was not imposed - it remained a princely state.

Defiers: Those units where the native ruler died without an heir, and British direct rule was not taken up. Those units where the native ruler died with an heir (or where there was no death of the ruler), and British direct rule was imposed.

(c) List the assumptions that are required to estimate a Complier average causal effect. In this context, which of those assumptions seem plausible (if any), and which seem suspect (if any)? Could you use any empirical methods to evaluate their plausibility?

There are 5 assumptions required to estimate a Complier average causal effect.

1. SUTVA (non-interference)
2. random assignment to treatment and control
3. potential outcomes are fixed attributes of each unit
4. exclusion restriction
5. no defier types
6. there are compliers, always takers, and never takers

SUTVA seems plausible. It is unlikely that the death of a ruler (without an heir) would affect the treatment status or outcome of a neighboring district. We could test this by a placebo test wherein we examine whether the death of an heirless ruler in a district affects the outcome of a neighboring district.

Random assignment of death of a ruler (without an heir) seems plausibly as-if random. Death seems like it strikes exogenously (aka not according to a clear pattern differentiated by districts). We could test this empirically with a balance test and an F statistic on pre-treatment covariates.

Potential outcomes seem to be a good way to think about this problem, and it is feasible to think about them as fixed attributes of a unit, as in the Neyman model. We can't test this.

Exclusion restriction seems broadly plausible. Yet, we could think of ways in which it could be challenged. For example, if princely states where there are no heirs (and the ruler dies) are those places in which mortality levels were higher (and thus the heirs died) because of bad development conditions, and that this could have long-term effects on the public goods outcomes of the district. We can't test this assumption, but have to reason it theoretically and substantively.

No defier types seems a reasonable assumption. This would mean that in places where the ruler died without an heir, the British didn't take over; and where there was no death or where there was an heir, the British did take over. We can't test this assumption.

It seems plausible to have compliers, always takers, and never takers. We can't test this assumption however.

(d) In some of her analyses, Iyer compares districts in which heirless rulers died during Dalhousie's rule to the remaining districts. Propose a design modification that could increase the plausibility of the assumptions you described in (c) and say how it does so.

A design modification could be to compare only the districts in which the ruler died (group 1: with heir; group 2: without heir), and exclude the districts where no death occurred. This might be sensible because districts where rulers are dying might be substantively different from those where they're not, in important ways that affect outcomes (like in mortality rates). As stated above, this could have challenges for the feasibility of the exclusion restriction. Thus, restricting the comparison only to districts with similar levels of deaths of rulers would increase our belief in this assumption.

Problem 3

Do legislators reward high municipal turnout with greater transfers? (i)

Problem 4

Problem 5

(a) The decline cannot readily be attributed to the effects of the new law. There are likely other events that happened during the time period that may affect the outcome of voting. Further, the law is not implemented as-if randomly across districts - it is implemented according to the level of poverty in an area. There could be important confounding variables with the level of poverty and voting.

(b) Intention to treat analysis.

```r
# There are 2000 units assigned to control, 75 percent have a value of 1
# There are 2000 units assigned to treatment, 75.25 % have a value of 1

treatment <- c(rep(0,2000), rep(1, 2000))

overall_outcome <- c(rep(1,1500), rep(0, 500), rep(1, 1505), rep(0, 495))

receipt <- c(rep(0,2000), rep(1,500), rep(0,1500))

data <- cbind(treatment, overall_outcome, receipt)

ITT <- function(treatment, outcome){

  mean_control <- mean(outcome[treatment == 0])
  mean_treatment <- mean(outcome[treatment == 1])
  itt <- mean_treatment - mean_control

  #standard error
  treated <- outcome[treatment == 1]
  var1 <- sum((treated - mean(treated))^2) / (length(treated) - 1)

  not_treated <- outcome[treatment == 0]
  var0 <- sum((not_treated - mean(not_treated))^2) / (length(not_treated) - 1)

  #se_itt
  se_itt <- sqrt((var1/length(treated)) + var0/length(not_treated))

  # t stat
  t_stat <- itt / se_itt

  df <- (var1/length(treated) + var0/length(not_treated))^2 /
((var1/length(treated))^2 / (length(treated) - 1) +
(var0/length(not_treated))^2 / (length(not_treated) - 1))

  # p value
  p_value <- pt(-abs(t_stat), df = df, ncp = 0, lower.tail = TRUE) +
    pt(abs(t_stat), df = df, ncp = 0, lower.tail = FALSE)

  return(c("ITT" = itt, "SE ITT" = se_itt, "p-value" = p_value))
}

itt <- ITT(treatment, overall_outcome)
```

```
print(itt)
```

```
##        ITT     SE ITT    p-value
## 0.00250000 0.01367353 0.85493672
```

We find that the intention to treatment analysis gives an estimate of -0.0025, and that the effect is not statistically different from 0. In other words, the effect of being assigned to the treatment is not statistically significant.

(c) Treatment receipt is "learning about the lower level of the fine." Thus, there is non-compliance in the experiment, shown in Table 1, where we see that there are people who are assigned to treatment, but did not receive the treatment (did not learn of the fine). The always-treats are those who are assigned to treatment and receive the treatment, and when assigned to control would still receive the treatment; we assume here that there aren't any, since we see only one-sided non compliance and since we would presume balancedness due to randomization since we don't see any noncompliance in the control group (thus no always takers) we assume there are none in the treatment. The never treats are those who when assigned to treatment receive the control, and when assigned to control receive the control; in this study these are the people assigned to treatment who did not receive the treatment (assuming no defiers). Finally, the compliers are those who when assigned to treatment receive the treatment, and when assigned to control receive the control; in this study, those in the treatment group who received the treatment could either be always-treats or compliers, and those in the control who received the control could be never-treats or compliers.

(d) Conduct IV analysis to estimate the complier average causal effect. Also, estimate the turnout (percentage voting) among Compliers in the control group.

```
## First, we will conduct IV analysis to estimate the complier
## average causal effect.

# The treatment assignment instruments for treatment receipt in the
# instrumental-variables estimator (wald estimator)

# in this case (with a binary treatment and a binary outcome)
# we can estimate the CACE as the same as the IV estimator
# It is Y^t - Y^c / X^t - X^c

# first, we get the numerator, which is the ITT that we already calculated

intention_to_treat <- 0.00250000

# now we need to get the denominator, which is the proportion of compliers

# first we estimate share of never-takers among those assigned to treatment
share_never_treated <- 1500/2000

# now we need to estimate the number of compliers among assigned to control
compliers_incontrol <- 2000 - (share_never_treated*2000)

# now we get the estimated share of compliers in the study
total_compliers <- compliers_incontrol + 500

share_compliers <- total_compliers/4000
```

```r
# now we calculate the CACE estimate
cace <- intention_to_treat/share_compliers

print(cace)
```

```
## [1] 0.01
```

```r
# now we can estimate the turnout rate among compliers assigned to the control
# group

# we have already estimated the number of compliers in the control group

compliers_incontrol
```

```
## [1] 500
```

```r
# we want the proportion of compliers in control
share_compliers_control <- compliers_incontrol/2000

# then we want the share of never treats in the treatment * their voting rate
voting_nevertreat <- share_never_treated*73

# now we can solve the equation for x, which is the estimate of turnout for
# the compliers in the control
turnout_compliers_control <- (75 - voting_nevertreat)/share_compliers_control
```

We estimate a CACE of 0.01. We also estimate the turnout among Compliers in the control group as 81%.

(e) Assumptions necessary for the IV analysis:

1) SUTVA (non-interference)
2) Random assignment to treatment and control
3) Potential outcomes are fixed attributes of each unit
4) Exclusion restriction
5) No defier types

Random assignment (2) seems plausible, given that the researchers experimentally manipulated the treatment. It seems plausible to assume that potential outcomes are fixed attributes of each unit (3). It seems fair to assume that there are no defier types (5) - after all, it is difficult to think of a unit that would seek out information if given the control, and ignore the information if given the treatment.

SUTVA (1) might face problems. It is possible that if people live closely together, the fact of someone in the treatment group receiving the treatment of information might lead this person to speak about it to someone assigned to control, thus affecting the outcome of the person in control who was not supposed to receive information. If the researchers can be certain that these units are not in contact, the threat of breaking SUTVA is less likely.

The exclusion restriction (4) can possibly be violated. This would be if assignment to treatment or control affects the outcome through another channel other than through the receipt of treatment (control). This could be the case if receiving information makes people more likely to tell their families to go vote because they will otherwise face a financial penalty, for example.

If either SUTVA or the exclusion restriction (or both) turn out not to be valid assumptions, then our analysis becomes less convincing, since it rests on these assumptions. With that said, it is possible to provide some evidence in support of SUTVA (or lack of interference) if the data is available - data on the outcomes of neighbors could be used here. The exclusion restriction would need to be substantiated with substantive and theoretical reasoning, there are no ready-made statistical tests to test its validity.

Problem 6