# Assignment 4

David Gyaraki, Thao Le

## Contents

```r
# load packages
if(!require(pacman)){install.packages("pacman")}

p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp,
       sampleSelection, quantreg, plm, nlme, knitr,car)
```

# 1 Question 1

| Judge | Jones | | Smith | |
|---|---|---|---|---|
| Sentences | Prison | Other | Prison | Other |
| Cases | 70% | 30% | 40% | 60% |
| Future arrests | 40% | 60% | 20% | 50% |

## 1.1 (i)

We can treat the following problem as follows: $Y_i$ is the outcome

## 1.2 (ii)

## 1.3 (iii)

# 2 Question 2

## 2.1 (i)

From what is given, we have $MDE = 0.1$, the power $p = 0.7$, the proportion of students in control group is $p = 0.5$. Assume the variance $\sigma^2 = 1$ To get the number of students the teacher should include in the experiment, we use the following formula:

$$\begin{aligned} n &= \left(\frac{t_{1-\alpha/2} - t_{1-q}}{MDE}\right)^2 \frac{\sigma^2}{p(1-p)} \\ &= \left(\frac{1.960 + 0.524}{0.1}\right)^2 \frac{1}{0.5(1-0.5)} \\ &\approx 2468 \end{aligned} \tag{1}$$

Thus, the teacher should include at least 2468 students in the experiment.

### 2.2 (ii)

This will change the proportion of students in treatment to $p = 0.5 \times 20\% = 0.1$, using the formula in Equation (1), the number of students required to participate in the experiment is:

$$\begin{aligned} n &= \left(\frac{1.960 + 0.524}{0.1}\right)^2 \frac{1}{0.1(1-0.9)} \\ &\approx 6856 \end{aligned} \tag{2}$$

Thus, the number of students required to participate in the experiment increases by $6856 - 2468 = 4388$ students.

### 2.3 (iii)

## 3 Question 3

### 3.1 (i)

```
# Load data
dfData = read.csv("AngristEvans80.csv")
attach(dfData)

# Fraction of girls among the first born child
count_girl1 = table(dfData$SEXK)
fraction_girl1= count_girl1[[2]]/(count_girl1[[1]]+count_girl1[[2]])

# Fraction of girls among the second born child
count_girl2 = table(dfData$SEX2ND)
fraction_girl2= count_girl2[[2]]/(count_girl2[[1]]+count_girl2[[2]])

cat("Fraction of girls among the first born child is:
 ↪  ",fraction_girl1,"\n", "Fraction of girls among the second born
 ↪  child is: ",fraction_girl2)
```

```
Fraction of girls among the first born child is:  0.4876463
 Fraction of girls among the second born child is:  0.4884266
```

```r
#Regress gender of second child on gender of first child
lm_second_first = lm(SEX2ND~SEXK, data = dfData)
summary(lm_second_first)
```

```
Call:
lm(formula = SEX2ND ~ SEXK, data = dfData)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4908 -0.4862 -0.4862  0.5092  0.5138

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.4861744  0.0008672 560.626   <2e-16 ***
SEXK        0.0046185  0.0012418   3.719    2e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4999 on 648470 degrees of freedom
Multiple R-squared:  2.133e-05, Adjusted R-squared:  1.979e-05
F-statistic: 13.83 on 1 and 648470 DF,  p-value: 2e-04
```

### 3.2 (ii)

```r
# First stage regression
lm_first_stage = lm(CHILD3 ~ SAMESEX, data= dfData)
summary(lm_first_stage)
```

```
Call:
lm(formula = CHILD3 ~ SAMESEX, data = dfData)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4093 -0.4093 -0.3552  0.5907  0.6448
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.3552366  0.0008544  415.79   <2e-16 ***
SAMESEX     0.0540534  0.0012051   44.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4852 on 648470 degrees of freedom
Multiple R-squared:  0.003093,  Adjusted R-squared:  0.003091
F-statistic:  2012 on 1 and 648470 DF,  p-value: < 2.2e-16
```

Is the instrumental variable sufficiently strong? => yes

```r
# Regress number of children on whether the first two children have the
↪  same gender
lm_total = lm(KIDCOUNT ~ SAMESEX, data= dfData)
summary(lm_total)
```

```
Call:
lm(formula = KIDCOUNT ~ SAMESEX, data = dfData)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5752 -0.5752 -0.5040  0.4248  9.4960

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.504033   0.001458 1716.93   <2e-16 ***
SAMESEX     0.071200   0.002057   34.61   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8283 on 648470 degrees of freedom
Multiple R-squared:  0.001844,  Adjusted R-squared:  0.001842
F-statistic:  1198 on 1 and 648470 DF,  p-value: < 2.2e-16
```

### 3.3 (iii)

In this study, the treatment group includes those who have a third child and the control group includes those who have two children or less. The variables that affect decision for mothers

to be assigned into treatment or control group is $Z = SAMESEX$, indicating whether the first two child are of the same sex or not.

The always takers are those who have a third child regardless of whether the first two children is of the same sex or not.

```
df_always = dfData[dfData$CHILD3 == 1,]
cat("The share of always takers is: ", nrow(df_always)/nrow(dfData))
```

```
The share of always takers is:  0.3824036
```

The compliers are those who only have a third child if the first two kids are of the same sex.

```
df_compliers1 = dfData[dfData$CHILD3 == 1 & dfData$SAMESEX == 1,]
df_compliers0 = dfData[dfData$CHILD3 == 0 & dfData$SAMESEX == 0,]
cat("The share of compliers is: ",
 ↪  (nrow(df_compliers1)+nrow(df_compliers0))/nrow(dfData))
```

```
The share of compliers is:  0.5264159
```

The never takers are those who will never have the third child regardless of whether the first two chidren are of the same sex or not.

```
df_never = dfData[dfData$CHILD3 == 0,]
cat("The share of never takers is: ", nrow(df_never)/nrow(dfData))
```

```
The share of never takers is:  0.6175964
```

Lastly, the defiers are those who will have a third child if the first two kids are of different sexes and will not have a third child if the first two kids are of the same sex.

```
df_defier1 = dfData[dfData$CHILD3 == 1 & dfData$SAMESEX == 0,]
df_defier0 = dfData[dfData$CHILD3 == 0 & dfData$SAMESEX == 1,]
cat("The share of defiers is: ",
 ↪  (nrow(df_defier1)+nrow(df_defier0))/nrow(dfData))
```

```
The share of defiers is:  0.4735841
```

**3.4 (iv)**

**3.5 (v)**

**3.6 (vi)**

**3.7 (vii)**