Problem Set 9

Econometrics

3a.

```
library(stats)
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
load("Angrist804049.Rda")


Angrist804049$YOB <- factor(Angrist804049$YOB)
Angrist804049$QOB <- factor(Angrist804049$QOB)


dummies <- model.matrix(~ QOB - 1, data = Angrist804049)
```

```r
dummies <- as.data.frame(dummies)
colnames(dummies) <- c("Q1", "Q2", "Q3", "Q4")
Angrist804049 <- cbind(Angrist804049, dummies)

dummies <- model.matrix(~ YOB - 1, data = Angrist804049)
dummies <- as.data.frame(dummies)
Angrist804049 <- cbind(Angrist804049, dummies)

head(Angrist804049)
```

```
##        AGE  AGEQ v3 EDUC ENOCENT ESOCENT v7        v8 LWKLYWGE MARRIED MIDATL
## 247200  47 47.00  2   12       0       0 14 10.197089 6.245846       1      0
## 247201  46 46.25  2   12       0       0 14  9.798405 5.847161       1      0
## 247203  47 47.00  2   16       0       0 18 10.657377 6.706133       1      0
## 247204  42 42.25  2   14       0       0 16 10.309119 6.357876       1      0
## 247205  44 44.25  2   12       0       0 14  9.393079 5.441835       1      0
## 247206  42 42.00  2   12       0       0 14 10.341904 6.390660       1      0
##        MT NEWENG v14 v15 CENSUS v17 QOB RACE SMSA SOATL v22 v23 WNOCENT WSOCENT
## 247200  0      0   0   1     80   1   1    1    0     0   2  52       0       0
## 247201  0      0   0   1     80  48   4    1    0     0   2  52       0       0
## 247203  0      0   0   1     80  22   1    1    0     0   2  52       0       0
## 247204  0      0   0   1     80  42   4    1    0     0   2  52       0       0
## 247205  0      0   0   1     80   5   4    1    0     0   2  52       0       0
## 247206  0      0   0   1     80  42   1    1    0     0   2  52       0       0
##        v26 YOB  AGEQ_SQ Q1 Q2 Q3 Q4 YOB30 YOB31 YOB32 YOB33 YOB34 YOB35 YOB36
## 247200   0  33 2209.000  1  0  0  0     0     0     0     1     0     0     0
## 247201   0  33 2139.062  0  0  0  1     0     0     0     1     0     0     0
## 247203   0  33 2209.000  1  0  0  0     0     0     0     1     0     0     0
## 247204   0  37 1785.062  0  0  0  1     0     0     0     0     0     0     0
## 247205   0  35 1958.062  0  0  0  1     0     0     0     0     0     1     0
```

```
## 247206   0  38 1764.000  1  0  0  0     0     0     0     0     0     0     0
##        YOB37 YOB38 YOB39 YOB40
## 247200     0     0     0     0
## 247201     0     0     0     0
## 247203     0     0     0     0
## 247204     1     0     0     0
## 247205     0     0     0     0
## 247206     0     1     0     0
```

3b.

```
iv_1 <- lm(EDUC ~ Q1 + Q2 + Q3, data = Angrist804049)
print(coeftest(iv_1, vcov = vcovHC(iv_1, type = "HC0")))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) 12.839431   0.011382 1128.0534 < 2.2e-16 ***
## Q1          -0.072497   0.016101   -4.5027 6.712e-06 ***
## Q2          -0.094716   0.016366   -5.7873 7.158e-09 ***
## Q3          -0.034018   0.015846   -2.1467   0.03181 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Being born in Q1 is associated with lower education levels.

```
yob_dummies <- paste0("YOB", unique(Angrist804049$YOB))
yob_dummies <- paste(yob_dummies, collapse = " + ")
formula <- as.formula(paste(
  "
  EDUC ~ Q1 + Q2 + Q3 + RACE + MARRIED + SMSA + NEWENG +
    MIDATL + ENOCENT + WNOCENT + SOATL + ESOCENT + WSOCENT + MT +
```

```
    AGE + AGEQ +
  ",
  yob_dummies
))

iv_2 <- lm(
  formula,
  data = Angrist804049
)

print(coeftest(iv_2, vcov = vcovHC(iv_2, type = "HC0")))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 16.43571000  0.21587930  76.1338 < 2.2e-16 ***
## Q1          -0.06717663  0.01592117  -4.2193 2.451e-05 ***
## Q2          -0.09140641  0.01589645  -5.7501 8.926e-09 ***
## Q3          -0.04040067  0.01541303  -2.6212  0.008762 **
## RACE        -1.56482985  0.02130454 -73.4505 < 2.2e-16 ***
## MARRIED      0.21970808  0.01699192  12.9302 < 2.2e-16 ***
## SMSA        -1.04637395  0.01458818 -71.7275 < 2.2e-16 ***
## NEWENG      -0.58482103  0.02788563 -20.9721 < 2.2e-16 ***
## MIDATL      -0.68112255  0.01990698 -34.2153 < 2.2e-16 ***
## ENOCENT     -1.06025422  0.01892923 -56.0115 < 2.2e-16 ***
## WNOCENT     -0.72469285  0.02433567 -29.7790 < 2.2e-16 ***
## SOATL       -1.09635926  0.02094968 -52.3330 < 2.2e-16 ***
## ESOCENT     -1.66124170  0.02830683 -58.6870 < 2.2e-16 ***
## WSOCENT     -0.93841557  0.02423990 -38.7137 < 2.2e-16 ***
```

```
## MT          -0.24503907  0.02831398  -8.6543 < 2.2e-16 ***
## AGE         -0.05963431  0.00473072 -12.6057 < 2.2e-16 ***
## YOB33        0.00564696  0.02491435   0.2267  0.820692
## YOB37       -0.01347382  0.02752471  -0.4895  0.624476
## YOB35       -0.00622840  0.02435304  -0.2558  0.798141
## YOB38        0.01269620  0.03008588   0.4220  0.673026
## YOB39       -0.00039777  0.03321564  -0.0120  0.990445
## YOB36       -0.02022262  0.02555575  -0.7913  0.428761
## YOB34       -0.02316179  0.02415337  -0.9589  0.337586
## YOB31        0.00659285  0.02822459   0.2336  0.815307
## YOB30       -0.03794129  0.03210385  -1.1818  0.237274
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, there is an association between being born in Q1 and education levels.

3c.

```
means <- aggregate(
  cbind(EDUC, LWKLYWGE) ~ Q1,
  data = Angrist804049, FUN = mean
)

diff_educ <- means$EDUC[1] - means$EDUC[2]
diff_wage <- means$LWKLYWGE[1] - means$LWKLYWGE[2]

result <- diff_wage / diff_educ

print(result)
```

```
## [1] 0.4433921
```

The Wald Estimator from the paper is 0.102. The result obstained here is higher at 0.443.

3d.

```
iv_3 <- ivreg(LWKLYWGE ~ EDUC | EDUC + Q1 + Q2 + Q3, data = Angrist804049)
summary(iv_3, vcov = vcovHC(iv_3, type = "HC0"))
```

```
##
## Call:
## ivreg(formula = LWKLYWGE ~ EDUC | EDUC + Q1 + Q2 + Q3, data = Angrist804049)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.75313 -0.23836  0.07355  0.33226  4.63850
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.9913306  0.0050866   981.3   <2e-16 ***
## EDUC        0.0709998  0.0003815   186.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6376 on 330448 degrees of freedom
## Multiple R-Squared: 0.1172,  Adjusted R-squared: 0.1172
## Wald test: 3.464e+04 on 1 and 330448 DF,  p-value: < 2.2e-16
```

Compared to the Wald estimator above at 0.443, the Iv estimator is lower at 0.071.

3e.

```
yob_dummies <- paste0("YOB", unique(Angrist804049$YOB))
interactions <- c()
for (yob in yob_dummies) {
  for (q in c("Q1", "Q2", "Q3")) {
    interactions <- c(interactions, paste0(yob, ":", q))
```

```
  }
}
interactions <- paste(interactions, collapse = " + ")

formula_string <- paste(
  "
  LWKLYWGE ~ EDUC + RACE + MARRIED + SMSA + NEWENG +
    MIDATL + ENOCENT + WNOCENT + SOATL + ESOCENT + WSOCENT + MT + AGE + AGEQ +
  ",
  yob_dummies,
  "| RACE + MARRIED + SMSA +
      NEWENG + MIDATL + ENOCENT + WNOCENT + SOATL +
      ESOCENT + WSOCENT + MT + AGE + AGEQ +",
  interactions,
  " + ",
  yob_dummies
)

iv_4 <- ivreg(
  formula_string,
  data = Angrist804049
)
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary_iv_4 <- summary(iv_4, vcov = vcovHC(iv_4, type = "HC0"))

print(paste("IV estimator: ", summary_iv_4$coef["EDUC", "Estimate"]))
```

```
## [1] "IV estimator:  0.108963872996728"
```

```
print(paste("IV standard error: ", summary_iv_4$coef["EDUC", "Std. Error"]))
```

```
## [1] "IV standard error:  0.024285890047 57"
```

The original paper has the Iv estimator at 0.0891 and the standard error at 0.0161. By caomparison, the coefficient obtained here is lower at 0.0634 and the standard error is significantly lower ay 0.000377.

3f.

```
ols <- lm(
  LWKLYWGE ~ EDUC + RACE + MARRIED + SMSA + NEWENG +
    MIDATL + ENOCENT + WNOCENT + SOATL + ESOCENT + WSOCENT + MT + AGE + AGEQ,
  data = Angrist804049
)

summary_ols <- summary(ols)

print(paste("OLS estimator: ", summary_ols$coef["EDUC", "Estimate"]))
```

```
## [1] "OLS estimator:  0.0634212451232428"
```

```
print(paste("OLS standard error: ", summary_ols$coef["EDUC", "Std. Error"]))
```

```
## [1] "OLS standard error:  0.000339592370276197"
```

3g.

```
formula_string <- paste(
  "
  LWKLYWGE ~ EDUC + RACE + MARRIED + SMSA + NEWENG +
    MIDATL + ENOCENT + WNOCENT + SOATL + ESOCENT + WSOCENT + MT + AGE + AGEQ +
  ",
  yob_dummies
)
```

```r
ols <- lm(
  formula_string,
  data = Angrist804049
)
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```r
iv_interval <- confint(
  iv_4, "EDUC",
  level = 0.95, vcov = vcovHC(iv_4, type = "HC0")
)
ols_interval <- confint(
  ols, "EDUC",
  level = 0.95, vcov = vcovHC(ols, type = "HC0")
)


print(paste("IV 95% confidence interval: ", iv_interval))
```

```
## [1] "IV 95% confidence interval:  0.0616904529708574"
## [2] "IV 95% confidence interval:  0.156237293022599"
```

```r
print(paste("OLS 95% confidence interval: ", ols_interval))
```

```
## [1] "OLS 95% confidence interval:  0.0627553532428805"
## [2] "OLS 95% confidence interval:  0.064086538550525"
```

The confidence intervals for OLS and IVs are very similar.