ECON 121 FA23 Problem Set 2

Robert Tso

Question 1

Verbal: list group members.

Robert Tso - A13829791

Tomas Lopez - A16215540

Stephanie Nguyen - A16798775

Akash Juwadi - A16372772

Question 2

Beta_0 is the intercept of the formula. Beta_1 is the coefficient that indicates the hourly wage return for each increase in level of education. Experience squared makes total experience a quadratic and looks to see if "excess" experience is a detriment to hourly wage, such as being too old or not having enough education.

Question 3

Code: Load packages and dataset, modify sample, generate variables, summarize data.

Verbal: Interpret the summary statistics.

The data's median is a 43 year old white male with 12 years of education, there is not much skew in the age and education. The max annual hours worked is over 4 times the median, and considering that the max and min weeks worked is only 2 apart

```
# The PDF will show the code you write here but not the output.
# Load packages and dataset, modify sample, generate variables here.
library(tidyverse)
library(fixest)
library(car)
```

```
# The PDF will show the code AND output here.
# Summarize the data here.
load("D:/Documents/Class/Econ 121/econ121/data/cps18.Rdata")
summary(cps18)
```

```
##       age             male           race             hrs_per_wk
##  Min.   :25.00   Min.   :0.000   Length:91928       Min.   :  0.00
##  1st Qu.:35.00   1st Qu.:0.000   Class :character   1st Qu.: 16.00
##  Median :44.00   Median :0.000   Mode  :character   Median : 40.00
##  Mean   :44.06   Mean   :0.478                      Mean   : 31.46
##  3rd Qu.:54.00   3rd Qu.:1.000                      3rd Qu.: 40.00
##  Max.   :64.00   Max.   :1.000                      Max.   :170.00
##                                                     NA's   :8877
##     wkswork         incwage           ed_lt_hs         ed_some_hs
##  Min.   : 0.00   Min.   :      0   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:26.00   1st Qu.:      0   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :52.00   Median :  30000   Median :0.00000   Median :0.00000
##  Mean   :38.57   Mean   :  43706   Mean   :0.03588   Mean   :0.06182
##  3rd Qu.:52.00   3rd Qu.:  60000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :52.00   Max.   :1609999   Max.   :1.00000   Max.   :1.00000
##
##   ed_hs_degree     ed_some_col      ed_ba_degree    ed_post_degree
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2739   Mean   :0.2711   Mean   :0.2282   Mean   :0.1291
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```

```
#drop rows with fewer than 50 weeks, worked fewer than 35 hours in a typical week,
#or has 0 dollars in annual earnings
cps <- cps18 %>%
  filter(
  wkswork>=50 &
  hrs_per_wk>=35 &
  incwage > 0
  )
```

```r
#calculate annual hours and hourly wage
annual_hours <- cps$wkswork*cps$hrs_per_wk
hr_wage <- cps$incwage/annual_hours

#add log hourly wage variable
 cps <- cps%>%
  mutate(
    log_hr_wage=log(hr_wage)
  )

#add race/ethnicity dummy variables
Black <- ifelse(cps$race == 'black', 1, 0)
Native <- ifelse(cps$race == 'native', 1, 0)
Asian <- ifelse(cps$race == 'asian/pacific', 1, 0)
Other <- ifelse(cps$race == 'multiple/other', 1, 0)

#add the dummy's to the cps dataframe
cps <- cps%>%
  mutate(
    Black=Black,
    Native=Native,
    Asian=Asian,
    Other=Other
  )

#add up the education columns with weights of the total years spent.
#since these are dummy variables, only one column for each row should be true
cps <- cps %>%
  mutate(education =
         ed_lt_hs*8 + #less than high school is assuming middle school completion
         ed_some_hs*10 + #some high school is 2 years or half for total 10
         ed_hs_degree*12 + #completing high school is total 12 years
         ed_some_col*14 + #some college is 2 years or half for total 14
         ed_ba_degree*16 + #BA is 16 years total
         ed_post_degree*20 #post degree is 2 years for 18 total
         )

#generate exper = age - education - 5
cps <- cps %>%
  mutate(exper = age - education - 5)

#generate exper2 = exper^2
cps <- cps %>%
  mutate(exper2 = exper^2)

#view(cps)
summary(cps)
```

```
##       age             male            race             hrs_per_wk
##  Min.   :25.00   Min.   :0.0000   Length:49153       Min.   : 35.00
##  1st Qu.:34.00   1st Qu.:0.0000   Class :character   1st Qu.: 40.00
##  Median :43.00   Median :1.0000   Mode  :character   Median : 40.00
##  Mean   :43.29   Mean   :0.5549                      Mean   : 43.37
```

```
##  3rd Qu.:52.00   3rd Qu.:1.0000                    3rd Qu.: 45.00
##  Max.   :64.00   Max.   :1.0000                    Max.   :170.00
##     wkswork         incwage           ed_lt_hs         ed_some_hs
##  Min.   :50.00   Min.   :      4   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:52.00   1st Qu.:  32000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :52.00   Median :  50000   Median :0.00000   Median :0.0000
##  Mean   :51.95   Mean   :  67129   Mean   :0.02382   Mean   :0.0401
##  3rd Qu.:52.00   3rd Qu.:  79000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :52.00   Max.   :1609999   Max.   :1.00000   Max.   :1.0000
##   ed_hs_degree     ed_some_col      ed_ba_degree    ed_post_degree
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2474   Mean   :0.2672   Mean   :0.2622   Mean   :0.1592
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##   log_hr_wage         Black            Native            Asian
##  Min.   :-6.254   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.: 2.691   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median : 3.111   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   : 3.125   Mean   :0.1177   Mean   :0.01229   Mean   :0.07804
##  3rd Qu.: 3.544   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   : 6.428   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##      Other          education         exper           exper2
##  Min.   :0.0000   Min.   : 8.00   Min.   : 0.00   Min.   :   0.0
##  1st Qu.:0.0000   1st Qu.:12.00   1st Qu.:14.00   1st Qu.: 196.0
##  Median :0.0000   Median :14.00   Median :23.00   Median : 529.0
##  Mean   :0.0179   Mean   :14.68   Mean   :23.61   Mean   : 681.1
##  3rd Qu.:0.0000   3rd Qu.:16.00   3rd Qu.:32.00   3rd Qu.:1024.0
##  Max.   :1.0000   Max.   :20.00   Max.   :51.00   Max.   :2601.0
```

Question 4

Code: Estimate regression.

Verbal: Interpret your results.

{log_hr_wage = 1.190526 + 0.106590*education + 0.027118*exper - 0.000397*exper2 }

Intercept only accounts for 1.190526 of the hourly wage, regardless of age, education or experience.Experience seems to have a relatively lower effect compared to education, but a reminder is that education is subtracted from age to get experience.Each year of Education has a 10.7% increase of hourly wage.

```
# All question 4 code here
# Define the Mincerian wage model

mince_model <- feols(log_hr_wage ~ education + exper + exper2, data = cps)

# Print the summary of the regression model
summary(mince_model)
```

```
## OLS estimation, Dep. Var.: log_hr_wage
## Observations: 49,153
## Standard-errors: IID
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  1.190526   0.021803  54.6039 < 2.2e-16 ***
## education    0.106590   0.001058 100.7750 < 2.2e-16 ***
## exper        0.027118   0.001199  22.6256 < 2.2e-16 ***
## exper2      -0.000397   0.000024 -16.5857 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.664179   Adj. R2: 0.175927
```

Question 5

Code: Estimate regression.

Verbal: Interpret your results.

log_hr_wage = 1.029361 + 0.109797*education* + *0.025147*exper - 0.000357*exper2* - *0.158205*black - 0.109256*Native* + *0.028961*Asian - 0.055523*Other* + *0.273182*male

Every year of Education is now 10.97% increase in wage, which is higher.

```
# All question 5 code here

ext_mince_model <- feols(log_hr_wage ~ education + exper + exper2
                         + Black + Native + Asian + Other
                         + male, data = cps)

# Print the summary of the regression model
summary(ext_mince_model)
```

```
## OLS estimation, Dep. Var.: log_hr_wage
## Observations: 49,153
## Standard-errors: IID
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)  1.029361   0.021838  47.13581  < 2.2e-16 ***
## education    0.109797   0.001042 105.33134  < 2.2e-16 ***
## exper        0.025147   0.001169  21.50449  < 2.2e-16 ***
## exper2      -0.000357   0.000023 -15.25868  < 2.2e-16 ***
## Black       -0.158205   0.009163 -17.26637  < 2.2e-16 ***
## Native      -0.109256   0.026574  -4.11139 3.9392e-05 ***
## Asian        0.028961   0.011033   2.62495 8.6690e-03 **
## Other       -0.055523   0.022088  -2.51369 1.1951e-02 *
## male         0.273182   0.005921  46.14068  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.647293   Adj. R2: 0.217218
```

Question 6

Code: Assess statistical significance.

Verbal: Interpret your results.

The Z-value is 40.904, which is really high, therefore it is statistically significant

```
# All question 6 code here

deltaMethod(ext_mince_model, "male - Black", rhs=0)
```

```
##              Estimate       SE    2.5 %   97.5 % Hypothesis z value  Pr(>|z|)
## male - Black 0.431387 0.010546 0.410716 0.452057   0.000000  40.904 < 2.2e-16
##
## male - Black ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 7

Code: Estimate regressions and assess significance of differences between them.

Verbal: Interpret your results.

The T-value for the two coefficients of male and female education is 2.300352, which is statistically significant.

```
# All question 7 code here

gender_mince_model <- feols(log_hr_wage ~ education + exper + exper2
                        + Black + Native + Asian + Other,
                        data = cps, split = ~male)

gender_mince_model
```

```
## Standard-errors: IID
## Sample: 0
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)  1.059452   0.031494  33.64004  < 2.2e-16 ***
## education    0.112598   0.001547  72.77250  < 2.2e-16 ***
## exper        0.017614   0.001640  10.73737  < 2.2e-16 ***
## exper2      -0.000214   0.000033  -6.43879 1.2293e-10 ***
## Black       -0.122881   0.012260 -10.02320  < 2.2e-16 ***
## Native      -0.111151   0.037439  -2.96884 2.9925e-03 **
## Asian        0.058325   0.015895   3.66932 2.4377e-04 ***
## Other       -0.035270   0.031434  -1.12205 2.6186e-01
## ---
## Sample: 1
##              Estimate Std. Error    t value  Pr(>|t|)
## (Intercept)  1.259626   0.029244  43.073669 < 2.2e-16 ***
## education    0.107783   0.001410  76.433081 < 2.2e-16 ***
## exper        0.032113   0.001658  19.373805 < 2.2e-16 ***
## exper2      -0.000485   0.000033 -14.810007 < 2.2e-16 ***
## Black       -0.194599   0.013596 -14.313114 < 2.2e-16 ***
## Native      -0.107007   0.037441  -2.858015 0.0042662 **
## Asian        0.008860   0.015270   0.580219 0.5617716
## Other       -0.070749   0.030849  -2.293394 0.0218329 *
```

```
#testing coefficients for education
(0.112598 - 0.107783)/sqrt((0.001547^2) + (0.001410^2))
```

```
## [1] 2.300352
```

Question 8

Code: Estimate regression, use delta method.

Verbal: Interpret your results.

The difference between the coefficients for male education and female education is 0.004815, which is exactly the coefficient of malexedu 0.004816.

Education, Experience, Exper2, and Black are the only ratios that are statistical significant from 1.

```
# All question 8 code here

#building the interaction terms
cps_interactions <- mutate(cps, malexedu = male*education,
                                malexexper = male*exper,
                                malexexper2 = male*exper2,
                                blackmale = male*Black,
                                nativemale = male*Native,
                                asianmale = male*Asian,
                                othermale = male*Other)

#building extended mincerian model with interaction terms
ext_gender_mince_model <- feols(log_hr_wage ~ education + exper + exper2
                         + Black + Native + Asian + Other + male
                         + malexedu + malexexper + malexexper2
                         + blackmale + nativemale
                         + asianmale + othermale,
                         data = cps_interactions)

summary(ext_gender_mince_model)
```

```
## OLS estimation, Dep. Var.: log_hr_wage
## Observations: 49,153
## Standard-errors: IID
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)  1.059452   0.032541 32.557028  < 2.2e-16 ***
## education    0.112598   0.001599 70.429645  < 2.2e-16 ***
## exper        0.017614   0.001695 10.391689  < 2.2e-16 ***
## exper2      -0.000214   0.000034 -6.231502 4.6572e-10 ***
## Black       -0.122881   0.012667 -9.700513  < 2.2e-16 ***
## Native      -0.111151   0.038685 -2.873256 4.0644e-03 **
## Asian        0.058325   0.016424  3.551194 3.8385e-04 ***
## Other       -0.035270   0.032479 -1.085922 2.7752e-01
## male         0.200174   0.043276  4.625565 3.7453e-06 ***
## malexedu    -0.004816   0.002109 -2.283249 2.2420e-02 *
## malexexper   0.014499   0.002343  6.189328 6.0897e-10 ***
## malexexper2 -0.000272   0.000047 -5.791825 7.0049e-09 ***
## blackmale   -0.071718   0.018340 -3.910390 9.2271e-05 ***
## nativemale   0.004144   0.053203  0.077889 9.3792e-01
## asianmale   -0.049465   0.022173 -2.230880 2.5694e-02 *
## othermale   -0.035479   0.044278 -0.801267 4.2298e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.646838   Adj. R2: 0.218205
```

```r
summary(gender_mince_model)
```

```
## Standard-errors: IID
## Sample: 0
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)  1.059452   0.031494  33.64004  < 2.2e-16 ***
## education    0.112598   0.001547  72.77250  < 2.2e-16 ***
## exper        0.017614   0.001640  10.73737  < 2.2e-16 ***
## exper2      -0.000214   0.000033  -6.43879 1.2293e-10 ***
## Black       -0.122881   0.012260 -10.02320  < 2.2e-16 ***
## Native      -0.111151   0.037439  -2.96884 2.9925e-03 **
## Asian        0.058325   0.015895   3.66932 2.4377e-04 ***
## Other       -0.035270   0.031434  -1.12205 2.6186e-01
## ---
## Sample: 1
##              Estimate Std. Error    t value  Pr(>|t|)
## (Intercept)  1.259626   0.029244  43.073669 < 2.2e-16 ***
## education    0.107783   0.001410  76.433081 < 2.2e-16 ***
## exper        0.032113   0.001658  19.373805 < 2.2e-16 ***
## exper2      -0.000485   0.000033 -14.810007 < 2.2e-16 ***
## Black       -0.194599   0.013596 -14.313114 < 2.2e-16 ***
## Native      -0.107007   0.037441  -2.858015 0.0042662 **
## Asian        0.008860   0.015270   0.580219 0.5617716
## Other       -0.070749   0.030849  -2.293394 0.0218329 *
```

```r
#extended mince model for
female_mince_model <- feols(log_hr_wage ~ education + exper + exper2
                        + Black + Native + Asian + Other,
                        data = cps,  subset = ~male==0)

male_mince_model <- feols(log_hr_wage ~ education + exper + exper2
                        + Black + Native + Asian + Other,
                        data = cps,  subset = ~male==1)

#deltaMethod test for each variable between male and female
deltaMethod(ext_gender_mince_model,"education/(education + malexedu)",RHS=1)
```

```
##                                  Estimate       SE    2.5 % 97.5 %
## education/(education + malexedu) 1.044679 0.019945 1.005588 1.0838
```

```r
deltaMethod(ext_gender_mince_model,"exper/(exper + malexexper)",RHS=1)
```

```
##                             Estimate       SE    2.5 % 97.5 %
## exper/(exper + malexexper) 0.548501 0.059572 0.431742 0.6653
```

```r
deltaMethod(ext_gender_mince_model,"exper2/(exper2 + malexexper2)",RHS=1)
```

```
##                               Estimate       SE    2.5 % 97.5 %
## exper2/(exper2 + malexexper2) 0.440431 0.076401 0.290689 0.5902
```

```
deltaMethod(ext_gender_mince_model,"Black/(Black + blackmale)",RHS=1)
```

```
##                             Estimate       SE    2.5 % 97.5 %
## Black/(Black + blackmale) 0.631455 0.078036 0.478508 0.7844
```

```
deltaMethod(ext_gender_mince_model,"Native/(Native + nativemale)",RHS=1)
```

```
##                               Estimate       SE    2.5 % 97.5 %
## Native/(Native + nativemale) 1.038726 0.506357 0.046284 2.0312
```

```
deltaMethod(ext_gender_mince_model,"Asian/(Asian + asianmale)",RHS=1)
```

```
##                             Estimate       SE    2.5 % 97.5 %
## Asian/(Asian + asianmale)   6.5831  11.2224 -15.4123 28.579
```

```
deltaMethod(ext_gender_mince_model,"Other/(Other + othermale)",RHS=1)
```

```
##                             Estimate       SE    2.5 % 97.5 %
## Other/(Other + othermale)  0.49853  0.50569 -0.49261 1.4897
```

```
#divide the coefficients directly to get a ratio, but no staistical significance.
coefficients(female_mince_model)/coefficients(male_mince_model)
```

```
## (Intercept)   education       exper      exper2       Black      Native
##   0.8410845   1.0446790   0.5485009   0.4404314   0.6314553   1.0387255
##       Asian       Other
##   6.5831134   0.4985263
```

Question 9

Code: Load dataset, compute unweighted and weighted means.

Verbal: Interpret your results.

The summary gives us the mean of black at 0.2502 and mean of hisp at 0.1578. When weighed black mean is 0.1386967 and hisp mean is 0.063047. When accounting for sampling weights, the estimates bias would be reduced.

```
# All question 9 code here

load("D:/Documents/Class/Econ 121/econ121/data/nlsy79.Rdata")
summary(nlsy79)
```

```
##      caseid        perweight           age79               region79
##  Min.   :    1   Min.   :   4100   Min.   :14.0   NORTHEAST    :2550
##  1st Qu.: 3172   1st Qu.:  88642   1st Qu.:16.0   NORTH CENTRAL:2934
##  Median : 6344   Median : 153970   Median :18.0   SOUTH        :4568
##  Mean   : 6344   Mean   : 264540   Mean   :17.9   WEST         :2395
##  3rd Qu.: 9515   3rd Qu.: 479563   3rd Qu.:20.0   NA's         : 239
##  Max.   :12686   Max.   :1058864   Max.   :22.0
##
##     foreign          urban14           mag14            news14
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:1.0000
##  Median :0.0000   Median :1.0000   Median :1.0000   Median :1.0000
##  Mean   :0.2198   Mean   :0.7884   Mean   :0.5627   Mean   :0.7611
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :5        NA's   :40       NA's   :85       NA's   :49
##      lib14           educ_mom        educ_dad        numsibs
##  Min.   :0.0000   Min.   : 0.00   Min.   : 0.00   Min.   : 0.000
##  1st Qu.:0.0000   1st Qu.: 9.00   1st Qu.: 9.00   1st Qu.: 2.000
##  Median :1.0000   Median :12.00   Median :12.00   Median : 3.000
##  Mean   :0.7079   Mean   :10.87   Mean   :10.95   Mean   : 3.854
##  3rd Qu.:1.0000   3rd Qu.:12.00   3rd Qu.:12.00   3rd Qu.: 5.000
##  Max.   :1.0000   Max.   :20.00   Max.   :20.00   Max.   :29.000
##  NA's   :53       NA's   :808     NA's   :1806    NA's   :18
##      afqt81        laborinc18        hours18          educ
##  Min.   : 1.00   Min.   :     0   Min.   :   0   Min.   : 0.00
##  1st Qu.:16.00   1st Qu.:     0   1st Qu.:   0   1st Qu.:12.00
##  Median :36.00   Median : 30000   Median :2040   Median :12.00
##  Mean   :40.91   Mean   : 44888   Mean   :1513   Mean   :13.11
##  3rd Qu.:65.00   3rd Qu.: 60000   3rd Qu.:2172   3rd Qu.:15.00
##  Max.   :99.00   Max.   :396970   Max.   :8736   Max.   :20.00
##  NA's   :808     NA's   :6115     NA's   :5866   NA's   :3
##      black             hisp             male
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.2502   Mean   :0.1578   Mean   :0.5047
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```

```r
#view(nlsy79)

weighted.mean(nlsy79$black,w=nlsy79$perweight)
```

```
## [1] 0.1386967
```

```r
weighted.mean(nlsy79$hisp,w=nlsy79$perweight)
```

```
## [1] 0.063047
```

Question 10

Code: Modify sample, estimate unweighted and weighted regressions.

Verbal: Assess unweighted versus weighted results.

When we account for the sampling weights, the magnitudes of all coefficients and intercepts EXCEPT for male, decrease, meanwhile male increases from 0.327012 to 0.390222. The swing of coefficients decreases from unweighted to weighted are marginal compared to the swing of coefficient increase for male.

We decided to stick with the unweighted dataset as the relatively drastic increase of male coefficient takes away from the importance of the other variables

```
# All question 10 code here

#drop rows with fewer than 50 weeks times 35 hours = 1750 annual hours worked,
#or has 0 dollars in annual earnings
nlsy <- nlsy79 %>%
  filter(
  hours18>=1750 &
  laborinc18 > 0
  )

#calculate hourly wage
hr_wage_79 <- nlsy$laborinc18/nlsy$hours18

#add log hourly wage variable
 nlsy <- nlsy%>%
  mutate(
    log_hr_wage=log(hr_wage_79)
  )

#generate exper = age - education - 5
nlsy <- nlsy %>%
  mutate(exper = age79 - educ - 5)

#generate exper2 = exper^2
nlsy <- nlsy %>%
  mutate(exper2 = exper^2)


summary(nlsy)
```

```
##      caseid        perweight           age79                region79
##  Min.   :    2   Min.   :   9471   Min.   :14.00   NORTHEAST    : 692
##  1st Qu.: 2363   1st Qu.: 122064   1st Qu.:16.00   NORTH CENTRAL: 939
##  Median : 4827   Median : 363376   Median :17.00   SOUTH        :1224
##  Mean   : 5203   Mean   : 330000   Mean   :17.39   WEST         : 666
##  3rd Qu.: 7746   3rd Qu.: 514266   3rd Qu.:19.00   NA's         :  49
##  Max.   :12663   Max.   :1058864   Max.   :22.00
##
##     foreign          urban14           mag14            news14
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:1.0000
##  Median :0.0000   Median :1.0000   Median :1.0000   Median :1.0000
```

```
##  Mean   :0.2382   Mean   :0.7886   Mean   :0.6062   Mean   :0.7796
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :1        NA's   :8        NA's   :17       NA's   :8
##      lib14          educ_mom       educ_dad        numsibs
##  Min.   :0.0000   Min.   : 0.0   Min.   : 0.00   Min.   : 0.000
##  1st Qu.:0.0000   1st Qu.:10.0   1st Qu.: 9.00   1st Qu.: 2.000
##  Median :1.0000   Median :12.0   Median :12.00   Median : 3.000
##  Mean   :0.7254   Mean   :11.1   Mean   :11.24   Mean   : 3.665
##  3rd Qu.:1.0000   3rd Qu.:12.0   3rd Qu.:13.00   3rd Qu.: 5.000
##  Max.   :1.0000   Max.   :20.0   Max.   :20.00   Max.   :17.000
##  NA's   :12       NA's   :168    NA's   :417     NA's   :5
##      afqt81        laborinc18        hours18          educ
##  Min.   : 1.00   Min.   :   173   Min.   :1750   Min.   : 0.00
##  1st Qu.:20.00   1st Qu.: 35000   1st Qu.:2080   1st Qu.:12.00
##  Median :42.00   Median : 54584   Median :2080   Median :13.00
##  Mean   :44.79   Mean   : 74179   Mean   :2370   Mean   :13.91
##  3rd Qu.:69.00   3rd Qu.: 85000   3rd Qu.:2600   3rd Qu.:16.00
##  Max.   :99.00   Max.   :396970   Max.   :8736   Max.   :20.00
##  NA's   :136
##      black             hisp             male         log_hr_wage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :-2.687
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 2.720
##  Median :0.0000   Median :0.0000   Median :1.0000   Median : 3.151
##  Mean   :0.2644   Mean   :0.1922   Mean   :0.5342   Mean   : 3.151
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.: 3.600
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   : 5.419
##
##      exper            exper2
##  Min.   :-11.000   Min.   :  0.00
##  1st Qu.: -4.000   1st Qu.:  1.00
##  Median : -1.000   Median :  4.00
##  Mean   : -1.513   Mean   : 14.32
##  3rd Qu.:  1.000   3rd Qu.: 16.00
##  Max.   : 16.000   Max.   :256.00
##
```

```
unweigh_nlsy = feols(log_hr_wage ~ educ + exper + exper2 + male + black + hisp, data = nlsy)
weighed_nlsy = feols(log_hr_wage ~ educ + exper + exper2 + male + black + hisp, data = nlsy, weights =

unweigh_nlsy
```

```
## OLS estimation, Dep. Var.: log_hr_wage
## Observations: 3,570
## Standard-errors: IID
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  1.373823   0.094204 14.58344 < 2.2e-16 ***
## educ         0.120598   0.006948 17.35798 < 2.2e-16 ***
## exper        0.007549   0.005396  1.39897 0.1619095
## exper2       0.001550   0.000658  2.35529 0.0185624 *
## male         0.327012   0.023025 14.20228 < 2.2e-16 ***
## black       -0.261476   0.027263 -9.59082 < 2.2e-16 ***
## hisp        -0.084562   0.030821 -2.74361 0.0061071 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.68309   Adj. R2: 0.224944
```

```
weighed_nlsy
```

```
## OLS estimation, Dep. Var.: log_hr_wage
## Observations: 3,570
## Weights: perweight
## Standard-errors: IID
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  1.358707   0.093546 14.52449  < 2.2e-16 ***
## educ         0.120667   0.006913 17.45457  < 2.2e-16 ***
## exper        0.005583   0.005437  1.02681 3.0458e-01
## exper2       0.000735   0.000720  1.02101 3.0732e-01
## male         0.390222   0.023060 16.92189  < 2.2e-16 ***
## black       -0.256713   0.035249 -7.28284 4.0033e-13 ***
## hisp        -0.074574   0.047890 -1.55718 1.1952e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 392.9   Adj. R2: 0.229852
```

Question 11

Verbal answer only, no code.

The coefficients for education and experience are different, CPS measures more impact from experience while nsly has more impact with education. Differences in coefficients should be expected as there are different explanatory variables that are not accounted in the mincerian equation from each respective dataset. For example nlsy parental data, sources of magazines, library, newspapers, which can supplement overall experience. There is also significant differences between the demographics of the datasets. The average age of CPS is 43 while for nsly it is 17, demonstrating a bias towards older and more established careers as someone who is 30+ may have more career opportunities than someone just graduating high school.

```
#comparing coefficients directly
coefficients(ext_mince_model)
```

```
##   (Intercept)     education         exper        exper2         Black
##  1.0293608948  0.1097965929  0.0251472070 -0.0003565336 -0.1582048751
##        Native         Asian         Other          male
## -0.1092556837  0.0289612040 -0.0555227677  0.2731816277
```

```
coefficients(unweigh_nlsy)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.373823027  0.120598455  0.007548522  0.001550007  0.327012316 -0.261476097
##         hisp
## -0.084562008
```

```
#comparing data summaries
summary(cps)
```

```
##       age             male            race             hrs_per_wk
##  Min.   :25.00   Min.   :0.0000   Length:49153       Min.   : 35.00
##  1st Qu.:34.00   1st Qu.:0.0000   Class :character   1st Qu.: 40.00
##  Median :43.00   Median :1.0000   Mode  :character   Median : 40.00
##  Mean   :43.29   Mean   :0.5549                      Mean   : 43.37
##  3rd Qu.:52.00   3rd Qu.:1.0000                      3rd Qu.: 45.00
##  Max.   :64.00   Max.   :1.0000                      Max.   :170.00
##     wkswork         incwage          ed_lt_hs        ed_some_hs
##  Min.   :50.00   Min.   :      4   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:52.00   1st Qu.:  32000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :52.00   Median :  50000   Median :0.00000   Median :0.0000
##  Mean   :51.95   Mean   :  67129   Mean   :0.02382   Mean   :0.0401
##  3rd Qu.:52.00   3rd Qu.:  79000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :52.00   Max.   :1609999   Max.   :1.00000   Max.   :1.0000
##   ed_hs_degree     ed_some_col      ed_ba_degree    ed_post_degree
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2474   Mean   :0.2672   Mean   :0.2622   Mean   :0.1592
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##   log_hr_wage         Black            Native            Asian
##  Min.   :-6.254   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
```

```
##  1st Qu.: 2.691   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median : 3.111   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   : 3.125   Mean   :0.1177   Mean   :0.01229   Mean   :0.07804
##  3rd Qu.: 3.544   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   : 6.428   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##      Other          education         exper           exper2
##  Min.   :0.0000   Min.   : 8.00   Min.   : 0.00   Min.   :   0.0
##  1st Qu.:0.0000   1st Qu.:12.00   1st Qu.:14.00   1st Qu.: 196.0
##  Median :0.0000   Median :14.00   Median :23.00   Median : 529.0
##  Mean   :0.0179   Mean   :14.68   Mean   :23.61   Mean   : 681.1
##  3rd Qu.:0.0000   3rd Qu.:16.00   3rd Qu.:32.00   3rd Qu.:1024.0
##  Max.   :1.0000   Max.   :20.00   Max.   :51.00   Max.   :2601.0
```

```
summary(nlsy)
```

```
##      caseid        perweight           age79                region79
##  Min.   :    2   Min.   :   9471   Min.   :14.00   NORTHEAST    : 692
##  1st Qu.: 2363   1st Qu.: 122064   1st Qu.:16.00   NORTH CENTRAL: 939
##  Median : 4827   Median : 363376   Median :17.00   SOUTH        :1224
##  Mean   : 5203   Mean   : 330000   Mean   :17.39   WEST         : 666
##  3rd Qu.: 7746   3rd Qu.: 514266   3rd Qu.:19.00   NA's         :  49
##  Max.   :12663   Max.   :1058864   Max.   :22.00
##
##     foreign          urban14           mag14            news14
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:1.0000
##  Median :0.0000   Median :1.0000   Median :1.0000   Median :1.0000
##  Mean   :0.2382   Mean   :0.7886   Mean   :0.6062   Mean   :0.7796
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :1        NA's   :8        NA's   :17       NA's   :8
##      lib14          educ_mom       educ_dad        numsibs
##  Min.   :0.0000   Min.   : 0.0   Min.   : 0.00   Min.   : 0.000
##  1st Qu.:0.0000   1st Qu.:10.0   1st Qu.: 9.00   1st Qu.: 2.000
##  Median :1.0000   Median :12.0   Median :12.00   Median : 3.000
##  Mean   :0.7254   Mean   :11.1   Mean   :11.24   Mean   : 3.665
##  3rd Qu.:1.0000   3rd Qu.:12.0   3rd Qu.:13.00   3rd Qu.: 5.000
##  Max.   :1.0000   Max.   :20.0   Max.   :20.00   Max.   :17.000
##  NA's   :12       NA's   :168    NA's   :417     NA's   :5
##      afqt81        laborinc18        hours18          educ
##  Min.   : 1.00   Min.   :   173   Min.   :1750   Min.   : 0.00
##  1st Qu.:20.00   1st Qu.: 35000   1st Qu.:2080   1st Qu.:12.00
##  Median :42.00   Median : 54584   Median :2080   Median :13.00
##  Mean   :44.79   Mean   : 74179   Mean   :2370   Mean   :13.91
##  3rd Qu.:69.00   3rd Qu.: 85000   3rd Qu.:2600   3rd Qu.:16.00
##  Max.   :99.00   Max.   :396970   Max.   :8736   Max.   :20.00
##  NA's   :136
##      black             hisp             male         log_hr_wage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :-2.687
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 2.720
##  Median :0.0000   Median :0.0000   Median :1.0000   Median : 3.151
##  Mean   :0.2644   Mean   :0.1922   Mean   :0.5342   Mean   : 3.151
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.: 3.600
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   : 5.419
```

```
##
##      exper              exper2
##  Min.   :-11.000   Min.   :  0.00
##  1st Qu.: -4.000   1st Qu.:  1.00
##  Median : -1.000   Median :  4.00
##  Mean   : -1.513   Mean   : 14.32
##  3rd Qu.:  1.000   3rd Qu.: 16.00
##  Max.   : 16.000   Max.   :256.00
##
```

Question 12

Verbal answer only, no code.

It is not a guarantee that education is a causal effect on hourly wages. As we have columns for childhood environment that affect a child's propensity for education, such as access to a library and educated parents. Any of these variables may possibly be the true causal effect that affect the education coefficient.

Question 13

Code: Estimate regression.

Verbal: Explain choice of covariates, interpret results.

At first glance AFQT may seem important as it measures job qualifications, however AFQT contains some overlaps with an IQ test, so it may just be a redundant approximation for education. We can observe that the education coefficient drops when accounting for afqt. Similarly the value of education from mom and dad have a positive impact at a marginal cost of the education coefficient, as parents that value education would probably impart that value onto their child, or at the very least force them to go to school regardless. When accounting for all 3 of these variables, education coefficient drops even more, indicating an importance to control AFQT as it is responsible for the largest swing.

```
# All question 13 code here

AFQT_nlsy = feols(log_hr_wage ~ educ + exper + exper2
                              + male + black + hisp
                              + afqt81,
                                data = nlsy)
```

```
## NOTE: 136 observations removed because of NA values (RHS: 136).
```

```
dad_nlsy = feols(log_hr_wage ~ educ + exper + exper2
                              + male + black + hisp
                              + educ_dad,
                                data = nlsy)
```

```
## NOTE: 417 observations removed because of NA values (RHS: 417).
```

```
mom_nlsy = feols(log_hr_wage ~ educ + exper + exper2
                              + male + black + hisp
                              + educ_mom,
                                data = nlsy)
```

```
## NOTE: 168 observations removed because of NA values (RHS: 168).
```

```
all_nsly = feols(log_hr_wage ~ educ + exper + exper2
                              + male + black + hisp
                              + afqt81 + educ_dad + educ_mom,
                                data = nlsy)
```

```
## NOTE: 617 observations removed because of NA values (RHS: 617).
```

```
coefficients(unweigh_nlsy)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.373823027  0.120598455  0.007548522  0.001550007  0.327012316 -0.261476097
##         hisp
## -0.084562008
```

```r
coefficients(AFQT_nlsy)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.743693692  0.062296790 -0.008806608  0.002105834  0.300097725 -0.050273498
##         hisp       afqt81
##  0.053087329  0.007691034
```

```r
coefficients(dad_nlsy)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.306690644  0.107281897  0.006545053  0.001683688  0.324705678 -0.213085772
##         hisp     educ_dad
##  0.002333010  0.021124570
```

```r
coefficients(mom_nlsy)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.160470217  0.110974242  0.006723204  0.001555521  0.327275878 -0.233805768
##         hisp     educ_mom
##  0.027571221  0.028922974
```

```r
coefficients(all_nsly)
```

```
##  (Intercept)         educ        exper       exper2         male        black
##  1.597785901  0.059188473 -0.009596746  0.002016371  0.316243618 -0.030427149
##         hisp       afqt81     educ_dad     educ_mom
##  0.120720557  0.006922848  0.010096062  0.008629175
```