HW_2
Shiqi Zhou

Exercise 1 OLS estimate
#1.1 Calculate the correlation between Y and X
```
> cor(Y.2009,X.2009)
       X_age.2009
[1,] NA -0.2204498
```

#1.2 Calculate the coefficients on this regression
```
> beta.2009
               Coef
intercept 24993.742
age        -229.538
```

#1.3 Calculate the standard errors

a) Using the standard formulas of the OLS
```
> standard.error.2009
 intercept     age2009
376.339545    7.239412
```

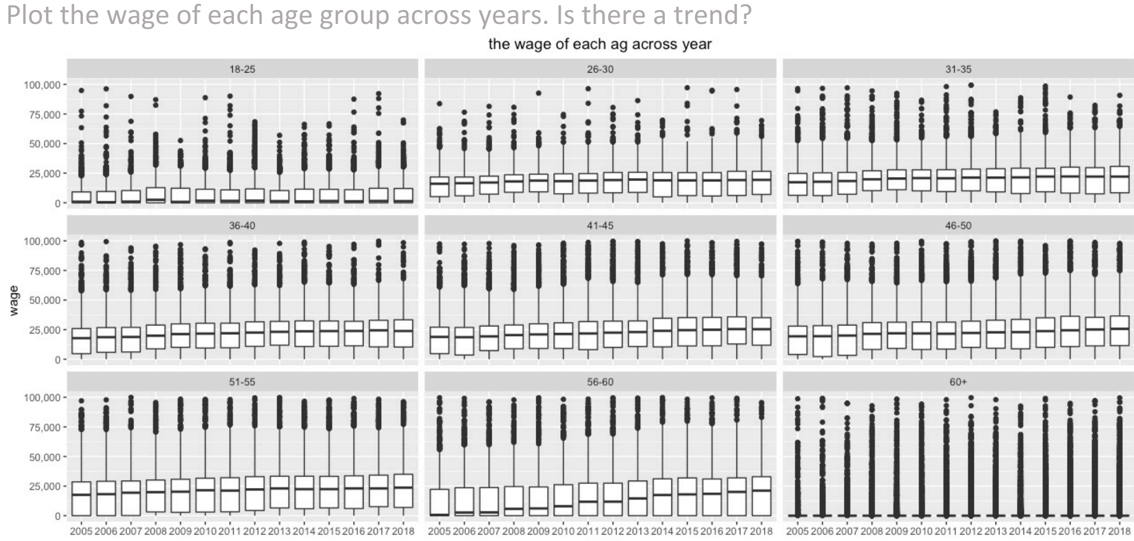
b) Using bootstrap with 49 and 499 replications respectively

Comment: with more replications, the results is closer with the OLS results.
```
> # use bootstrap with 49 replications
> bootstap(49,a,"wage","age")
          Coefficients Standard.error
Intercept 24937.26     278.4001
Age       -228.9561    4.999106
> # use bootstrap with 499 replications
> bootstap(499,a,"wage","age")
          Coefficients Standard.error
Intercept 24999.34     320.714
Age       -229.5425    5.360148
```

Exercise 2 Detrend Data
#2.1 Create a categorical variable ag

In code, Left 281858 observations used in following questions

#2.2 Plot the wage of each age group across years. Is there a trend?



Comment:

1) the average wage of “18-25”, “26-30”,”31-35” across year doesn’t seem to change a lot.
2) There is a increasing trend across year in the average wage of “36-40”, “41-45”,”46-50”, “51-55”, and “56-60”, and the trend is especially obvious in “56-60”

#2.3 After including a time fixed effect, how do the estimated coefficients change?

Comment:

Coefficients doesn’t change a lot.

```
> beta_year
                  [,1]
V1         23701.21149
age         -239.43116
yeary2006     15.35177
yeary2007    325.86790
yeary2008   1502.39383
yeary2009   1773.88915
yeary2010   1919.33792
yeary2011   2178.45694
yeary2012   2687.65210
yeary2013   2629.52902
yeary2014   2897.29765
yeary2015   3302.41900
yeary2016   3621.13249
yeary2017   3683.36204
yeary2018   3895.74385
```

 # Exercise 3 Numerical Optimization

#3.1 Exclude all individuals who are inactive

Left 11540 observations used in following questions

#3.2 Write a function that returns the likelihood of the probit of being employed

Functions in Code.

#3.3 Optimize the model and interpret the coefficients

```
> (rr2 =filter(rr1,min.nega==min(min.nega)))
  Coef.Con    Coef.age min.nega
1 1.045492 0.006893732  3555.89
```

#=============interpret=====================

with several times searching, find the minimum negative likelihood, which is the maximum likelihood,

as results, the coeffi.con is 1.045492, the coeffi.age is 0.006893732, and the min negative likelihood is 3555.89

which is checked by logLik() above.

from result, we find that, age has a positive effect on labor market participation,

in specific, one year older increases the probability of labor participation by 0.006893732

#3.4 Can you estimate the same model including wages as a determinant of labor market participation? Explain.

Answer: cannot use same model including wage.

Explanation:

Wage and empstat show the same massage whether this person is working or not.
If someone is Unemployed, his/her wage will be zero.
So, wage will be omitted with the empstat.
Thus, cannot use the same method.

- warning when using glm, it is not suitable for this method

```
> reg3 = glm(y.empwage~x.age+x.wage,family = binomial(link = "probit"))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> beta_emp <-summary(reg3)$coefficients[c("(Intercept)", "x.age","x.wage"),c("Estimate")]
> beta_emp
 (Intercept)        x.age       x.wage
4.644814e-01 1.780770e-03 6.058113e-05
```

- with right likelihood function, the optimization results is different with the glm() results

  glm() results

```
> likelihood2(beta_emp,y.empwage,x.age,x.wage)
[1] 2795.405
> logLik(reg3) #return -likelihood(beta_emp,y,x)
'log Lik.' -2833.345 (df=3)
```

  optimization results

```
> (rr4 =filter(rr3,min.nega==min(min.nega)))
   Coef.Con   Coef.age     Coef.wage min.nega
1 -1.514389 0.03916903 0.0001717977 3582.559
```

 # Exercise 4 Discrete choice

#4.1 Exclude all individuals who are inactive

In code, left 128636 observations used in following questions.

#4.2 Write and optimize the probit, logit, and the linear probability models

"min.nega" is the negative value of likelihood, so the minimum of it is the maximum of likelihood

a) Results of probit (results checked with glm())

```
> (probitrr =filter(rr_year,min.nega==min(min.nega)))
   Coef.Con   Coef.age Coef.yr2006 Coef.yr2007 Coef.yr2008 Coef.yr2009 Coef.yr2010 Coef.yr2011 Coef.yr2012 Coef.yr2013 Coef.yr2014 Coef.yr2015 min.nega
1 0.7295886 0.01281722  0.01716061    8.000366   0.1099654  0.02668693  0.02185919   0.0551456  0.01007445 -0.04008155 -0.03371167 -0.05380878 51021.82
```

b) Results of logit (results checked with glm())

```
> (logitrr =filter(rr_year2,min.nega==min(min.nega)))
  Coef.Con   Coef.age Coef.yr2006 Coef.yr2007 Coef.yr2008 Coef.yr2009 Coef.yr2010 Coef.yr2011 Coef.yr2012 Coef.yr2013 Coef.yr2014 Coef.yr2015 min.nega
1 1.120039 0.02531756  0.03170437   0.1571909   0.2127451  0.04556066  0.03729911   0.1008563  0.01197203 -0.08518204 -0.07183874   -0.111437 42213.76
```

c) Results of linear (results is different in own function and lm())

Use own function, get this result

```
> (linearrr =filter(rr_year3,min.nega==min(min.nega)))
  Coef.Con  Coef.age Coef.yr2006 Coef.yr2007 Coef.yr2008 Coef.yr2009 Coef.yr2010 Coef.yr2011 Coef.yr2012 Coef.yr2013 Coef.yr2014 Coef.yr2015 min.nega
1 1.357489 0.0261681    -1.10654   0.6604134  -0.9725123   -0.780083    1.244234  -0.0112769   -1.826075   -0.713471  -0.2200551   0.3210002 143512.7
```

Directly use lm()

```
> beta_emp
  (Intercept)    x.year.age        yr2006        yr2007        yr2008        yr2009        yr2010        yr2011        yr2012        yr2013        yr2014
 0.7977483650  0.0023358617  0.0029332879  0.0139479286  0.0184425587  0.0040834132  0.0033035718  0.0088873932  0.0008988494 -0.0083476676 -0.0070497876
```

Hessian matrices in optimizing are always 0, so, not report here

#4.3 Interpret and compare the estimated coefficients. How significant are they?

Coefficients:

1) In probit model, Coef.age is 0.01281, Coef.con is 0.7296
2) In logit model, Coef.age is 0.02531, Coef.con is 1.1200
3) In linear model, Coef.age is 0.0261681, Coef.con is 1.357489
4) Coefficients in probit and logit model relatively closer, compared with linear model. And the coefficient of age in probit is the smallest. All age in different models has positive effects on

participation in labor market, age increases 1, the probability of participating labor market increases 0.01281, 0.02531, 0.0261681 in probit, logit and linear model, respectively.

Significance: Hessian matrices in optimizing are always 0, so, do not know the significance

Exercise 5 Marginal Effects

#5.1 Compute the marginal effect of the previous probit and logit models

a) Probit margin

```
> probit_margin
             probit_margin
Coef.Con       0.118865532
Coef.age       0.002088199
Coef.yr2006    0.002795829
Coef.yr2007    1.303430185
Coef.yr2008    0.017915712
Coef.yr2009    0.004347870
Coef.yr2010    0.003561327
Coef.yr2011    0.008984393
Coef.yr2012    0.001641342
Coef.yr2013   -0.006530138
Coef.yr2014   -0.005492349
Coef.yr2015   -0.008766597
```

b) Logit margin

```
> logit_margin
             logit_margin
Coef.Con      0.102897783
Coef.age      0.002325919
Coef.yr2006   0.002912674
Coef.yr2007   0.014441100
Coef.yr2008   0.019544845
Coef.yr2009   0.004185648
Coef.yr2010   0.003426662
Coef.yr2011   0.009265652
Coef.yr2012   0.001099868
Coef.yr2013  -0.007825657
Coef.yr2014  -0.006599811
Coef.yr2015  -0.010237697
```

#5.2 Construct the standard errors of the marginal effects

a) Probit se

```
> seprob
             Con          age     yr2006     yr2007     yr2008     yr2009     yr2010     yr2011     yr2012     yr2013     yr2014     yr2015
Con    0.0351556          NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN
age          NaN 0.0006227443        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN
yr2006       NaN          NaN 0.03432725 0.02440238 0.02440225 0.02440250 0.02440289 0.02440316 0.02440346 0.02440372 0.02440436 0.02440470
yr2007       NaN          NaN 0.02440238 0.03402211 0.02440227 0.02440258 0.02440303 0.02440335 0.02440370 0.02440401 0.02440476 0.02440516
yr2008       NaN          NaN 0.02440225 0.02440227 0.03408462 0.02440235 0.02440260 0.02440277 0.02440296 0.02440312 0.02440353 0.02440374
yr2009       NaN          NaN 0.02440250 0.02440258 0.02440235 0.03409594 0.02440343 0.02440388 0.02440439 0.02440482 0.02440590 0.02440646
yr2010       NaN          NaN 0.02440289 0.02440303 0.02440260 0.02440343 0.03381311 0.02440554 0.02440652 0.02440734 0.02440942 0.02441051
yr2011       NaN          NaN 0.02440316 0.02440335 0.02440277 0.02440388 0.02440554 0.03361059 0.02440802 0.02440913 0.02441192 0.02441338
yr2012       NaN          NaN 0.02440346 0.02440370 0.02440296 0.02440439 0.02440652 0.02440802 0.03317592 0.02441113 0.02441471 0.02441658
yr2013       NaN          NaN 0.02440372 0.02440401 0.02440312 0.02440482 0.02440734 0.02440913 0.02441113 0.03390180 0.02441706 0.02441928
yr2014       NaN          NaN 0.02440436 0.02440476 0.02440353 0.02440590 0.02440942 0.02441192 0.02441471 0.02441706 0.03379227 0.02442608
yr2015       NaN          NaN 0.02440470 0.02440516 0.02440374 0.02440646 0.02441051 0.02441338 0.02441658 0.02441928 0.02442608 0.03390140
```

b) Logit se

```
> selog
             Con          age     yr2006     yr2007     yr2008     yr2009     yr2010     yr2011     yr2012     yr2013     yr2014     yr2015
Con    0.0351556          NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN
age          NaN 0.0006227443        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN        NaN
yr2006       NaN          NaN 0.03432725 0.02440238 0.02440225 0.02440250 0.02440289 0.02440316 0.02440346 0.02440372 0.02440436 0.02440470
yr2007       NaN          NaN 0.02440238 0.03402211 0.02440227 0.02440258 0.02440303 0.02440335 0.02440370 0.02440401 0.02440476 0.02440516
yr2008       NaN          NaN 0.02440225 0.02440227 0.03408462 0.02440235 0.02440260 0.02440277 0.02440296 0.02440312 0.02440353 0.02440374
yr2009       NaN          NaN 0.02440250 0.02440258 0.02440235 0.03409594 0.02440343 0.02440388 0.02440439 0.02440482 0.02440590 0.02440646
yr2010       NaN          NaN 0.02440289 0.02440303 0.02440260 0.02440343 0.03381311 0.02440554 0.02440652 0.02440734 0.02440942 0.02441051
yr2011       NaN          NaN 0.02440316 0.02440335 0.02440277 0.02440388 0.02440554 0.03361059 0.02440802 0.02440913 0.02441192 0.02441338
yr2012       NaN          NaN 0.02440346 0.02440370 0.02440296 0.02440439 0.02440652 0.02440802 0.03317592 0.02441113 0.02441471 0.02441658
yr2013       NaN          NaN 0.02440372 0.02440401 0.02440312 0.02440482 0.02440734 0.02440913 0.02441113 0.03390180 0.02441706 0.02441928
yr2014       NaN          NaN 0.02440436 0.02440476 0.02440353 0.02440590 0.02440942 0.02441192 0.02441471 0.02441706 0.03379227 0.02442608
yr2015       NaN          NaN 0.02440470 0.02440516 0.02440374 0.02440646 0.02441051 0.02441338 0.02441658 0.02441928 0.02442608 0.03390140
```