Econ 613 Homework 2

Lu Liu

Exercise 1 OLS estimate

1. Calculate the correlation between Y and X

0.143492

2. Calculate the coefficients on this regression. remember, $\hat{\beta} = (X^T X)^{-1} X^T Y$

beta_hat: 230.9923

3. Calculate the standard errors of β

- Using the standard formulas of the OLS.

se_betah :14.8774

- Using bootstrap with 49 and 499 replications respectively. Comment on the difference between the two strategies.
- When R = 49

```
> estimate
                                    estmean     estsd
(Intercept) 14853.9230 635.46022 14149.8497 596.08216
age           213.4264  14.63047   232.1823  15.72919
```

- When R = 499

```
> estimate2
                                   estmean2    estsd2
(Intercept) 14204.9097 629.95432 14100.5044 605.43018
age           234.0999  14.52747   232.3819  16.16779
```

According to the result, the standard error in the first strategy is smaller than the rest in the second strategy. The standard formula of the OLS is more common while Bootstrapping is more important in smaller sample, and it allows the repeated observations.

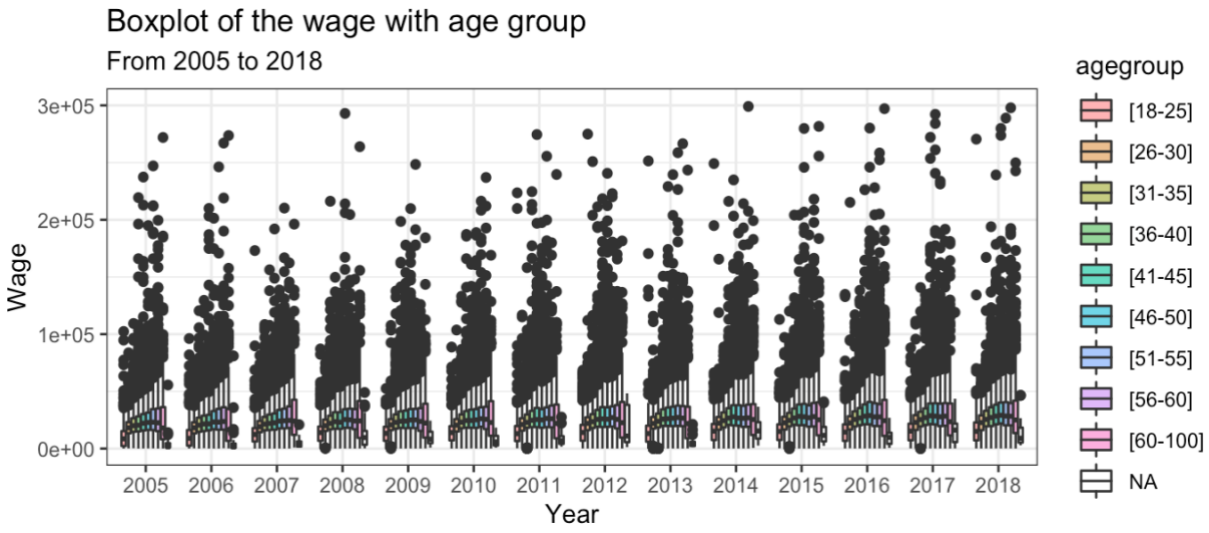
Exercise 2 Detrend Data

1. Create a categorical variable ag, which bins the age variables into the following groups: “18-25”, “26- 30”, “31-35”, “36-40”, “41-45”, “46-50”, “51-55”, “56-60”, and “60+”.

Screenshot of parts of the data.

	X	idind	idmen	year	empstat	respondent	profession	gender	age	wage	agegroup
1	1	1120001004058009984	1200010040580100	2005	Inactive	1		Female	31	12334	[31-35]
3	3	1120001006663010048	1200010066630100	2005	Employed	1	38	Male	32	50659	[31-35]
4	4	1120001006663010048	1200010066630100	2005	Employed	0	45	Female	28	19231	[26-30]
6	6	1120001008644009984	1200010086440100	2005	Employed	1	34	Male	37	31511	[36-40]
7	7	1120001008644009984	1200010086440100	2005	Employed	0	42	Female	35	24873	[31-35]
8	8	1120001010299010048	1200010102990100	2005	Employed	1	55	Female	41	30080	[41-45]
10	10	1120001011845010048	1200010118450100	2005	Employed	1	37	Male	55	43296	[51-55]
11	11	1120001011845010048	1200010118450100	2005	Employed	0	54	Female	55	20426	[51-55]
19	19	1120002004513010048	1200020045130100	2005	Employed	1	34	Female	55	49240	[51-55]
21	21	1120002009437010048	1200020094370100	2005	Employed	1	22	Male	41	15005	[41-45]
22	22	1120002009437010048	1200020094370100	2005	Employed	0	34	Female	39	35192	[36-40]
26	26	1120002012268009984	1200020122680100	2005	Employed	1	45	Male	30	22852	[26-30]
27	27	1120002012268009984	1200020122680100	2005	Inactive	0		Female	32	1832	[31-35]
29	29	1120014901293010048	1200149012930100	2005	Employed	1	46	Male	42	28247	[41-45]
30	30	1120014901293010048	1200149012930100	2005	Employed	0	55	Female	36	21134	[36-40]
39	39	1120017701262009984	1200177012620100	2005	Employed	1	48	Male	56	21051	[56-60]
40	40	1120017701262009984	1200177012620100	2005	Employed	0	54	Female	54	16168	[51-55]
41	41	1120017701262009984	1200177012620101	2005	Employed	0	63	Male	27	19688	[26-30]
42	42	1120017701262009984	1200177012620102	2005	Employed	0	54	Female	31	11666	[31-35]

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



Yes, the wage is increasing from 2005 to 2018 for almost all groups.

3. Consider $Y_{it} = \beta X_{it} + \gamma_t + e_{it}$. After including a time fixed effect, how do the estimated coefficients change?

```
Call:
lm(formula = wage ~ age, data = ag)

Residuals:
    Min      1Q  Median      3Q     Max
 -41113  -10548   -2616    5795 1379563

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 9420.272    214.285   43.96   <2e-16 ***
age          413.789      4.966   83.33   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20990 on 138028 degrees of freedom
Multiple R-squared:  0.0479,    Adjusted R-squared:  0.04789
F-statistic:  6944 on 1 and 138028 DF,  p-value: < 2.2e-16
```

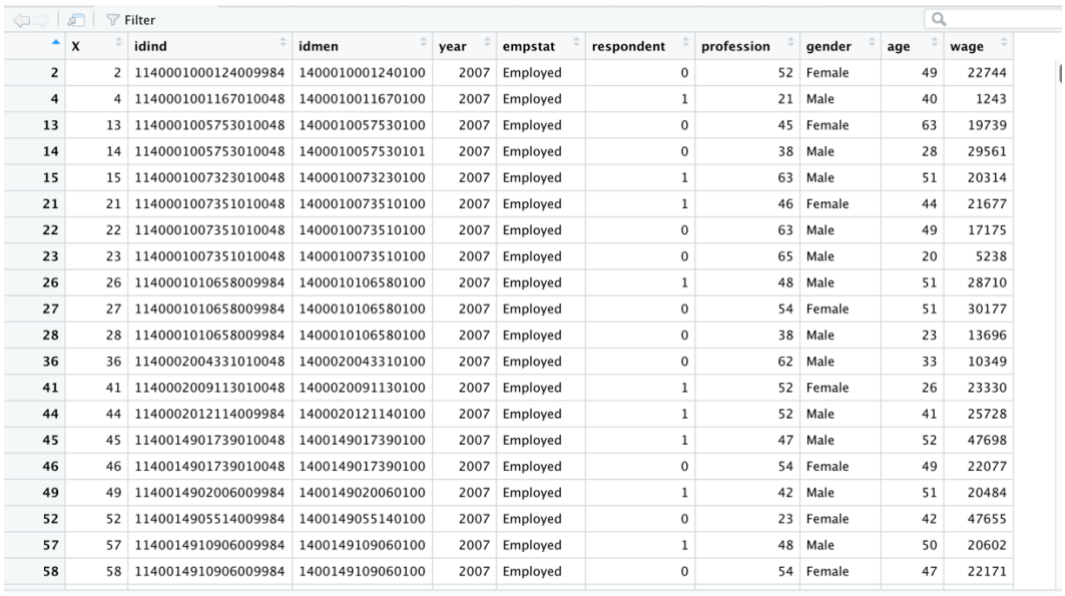
```
Call:
lm(formula = wage ~ age + year, data = ag)

Residuals:
    Min      1Q  Median      3Q     Max
 -38975  -10350   -2578    5678 1379564

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6559.741    287.832  22.790  < 2e-16 ***
age          397.989      4.967  80.135  < 2e-16 ***
year2006      76.694    293.215   0.262    0.794
year2007    1335.278    297.769   4.484 7.32e-06 ***
year2008    2546.152    299.111   8.512  < 2e-16 ***
year2009    2728.219    296.626   9.198  < 2e-16 ***
year2010    3122.737    295.836  10.556  < 2e-16 ***
year2011    3554.625    292.813  12.140  < 2e-16 ***
year2012    4262.788    289.308  14.734  < 2e-16 ***
year2013    4278.128    297.464  14.382  < 2e-16 ***
year2014    4809.050    296.479  16.221  < 2e-16 ***
year2015    5284.613    297.816  17.745  < 2e-16 ***
year2016    5764.943    297.330  19.389  < 2e-16 ***
year2017    5821.031    301.327  19.318  < 2e-16 ***
year2018    6142.024    304.010  20.203  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20900 on 138015 degrees of freedom
Multiple R-squared:  0.0561,    Adjusted R-squared:  0.056
F-statistic: 585.9 on 14 and 138015 DF,  p-value: < 2.2e-16
```

The estimated coefficient of age decreases after including a time fixed effect.

Exercise 3 Numerical Optimizations

1. Exclude all individuals who are inactive.

Screenshot of parts of the data.



Filter

	X	idind	idmen	year	empstat	respondent	profession	gender	age	wage
2	2	1140001000124009984	1400010001240100	2007	Employed	0	52	Female	49	22744
4	4	1140001001167010048	1400010011670100	2007	Employed	1	21	Male	40	1243
13	13	1140001005753010048	1400010057530100	2007	Employed	0	45	Female	63	19739
14	14	1140001005753010048	1400010057530101	2007	Employed	0	38	Male	28	29561
15	15	1140001007323010048	1400010073230100	2007	Employed	1	63	Male	51	20314
21	21	1140001007351010048	1400010073510100	2007	Employed	1	46	Female	44	21677
22	22	1140001007351010048	1400010073510100	2007	Employed	0	63	Male	49	17175
23	23	1140001007351010048	1400010073510100	2007	Employed	0	65	Male	20	5238
26	26	1140001010658009984	1400010106580100	2007	Employed	1	48	Male	51	28710
27	27	1140001010658009984	1400010106580100	2007	Employed	0	54	Female	51	30177
28	28	1140001010658009984	1400010106580100	2007	Employed	0	38	Male	23	13696
36	36	1140002004331010048	1400020043310100	2007	Employed	0	62	Male	33	10349
41	41	1140002009113010048	1400020091130100	2007	Employed	1	52	Female	26	23330
44	44	1140002012114009984	1400020121140100	2007	Employed	1	52	Male	41	25728
45	45	1140014901739010048	1400149017390100	2007	Employed	1	47	Male	52	47698
46	46	1140014901739010048	1400149017390100	2007	Employed	0	54	Female	49	22077
49	49	1140014902006009984	1400149020060100	2007	Employed	1	42	Male	51	20484
52	52	1140014905514009984	1400149055140100	2007	Employed	0	23	Female	42	47655
57	57	1140014910906009984	1400149109060100	2007	Employed	1	48	Male	50	20602
58	58	1140014910906009984	1400149109060100	2007	Employed	0	54	Female	47	22171

2. Write a function that returns the likelihood of the probit of being employed.

```
fnlike<- function(par, age, empstat)
{
  xb = par[1] + par[2]*age
  p = pnorm(xb)
  p[p>0.999999] = 0.999999
  p[p<0.000001] = 0.000001
  like = empstat*log(p) +(1-empstat)*log(1-p)
  return(-sum(like))
}
```

3. Optimize the model and interpret the coefficients. You can use pre-programmed optimization packages.

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

No, we cannot estimate the same model including wages as a determinant of labor market participation.

Exercise 4 Discrete choice

1. Exclude all individuals who are inactive.

Screenshot of parts of the data.

	X	idind	idmen	year	empstat	respondent	profession	gender	age	wage
3	3	1120001006663010048	1200010066630100	2005	Employed	1	38	Male	32	50659
4	4	1120001006663010048	1200010066630100	2005	Employed	0	45	Female	28	19231
6	6	1120001008644009984	1200010086440100	2005	Employed	1	34	Male	37	31511
7	7	1120001008644009984	1200010086440100	2005	Employed	0	42	Female	35	24873
8	8	1120001010299010048	1200010102990100	2005	Employed	1	55	Female	41	30080
10	10	1120001011845010048	1200010118450100	2005	Employed	1	37	Male	55	43296
11	11	1120001011845010048	1200010118450100	2005	Employed	0	54	Female	55	20426
19	19	1120002004513010048	1200020045130100	2005	Employed	1	34	Female	55	49240
21	21	1120002009437010048	1200020094370100	2005	Employed	1	22	Male	41	15005
22	22	1120002009437010048	1200020094370100	2005	Employed	0	34	Female	39	35192
26	26	1120002012268009984	1200020122680100	2005	Employed	1	45	Male	30	22852
29	29	1120014901293010048	1200149012930100	2005	Employed	1	46	Male	42	28247
30	30	1120014901293010048	1200149012930100	2005	Employed	0	55	Female	36	21134

Showing 1 to 13 of 106,538 entries, 10 total columns

2. Write and optimize the probit, logit, and the linear probability models.