Exercise 1 OLS estimate

Q1：Calculate the correlation between Y and X.

-0.179

Q2：Calculate the coefficients on this regression

The intercept is 22075.1066

The coefficient of age is -180.1765

Q3: Calculate the standard errors

1. **Using the standard formulas of the OLS.**
   Standard error of intercept: 357.83
   Standard error of coefficient of age is 6.97
2. **Using bootstrap with 49 and 499 replications respectively. Comment on the difference between the two strategies.**
   Bootstrap with 49:
   Standard error of intercept: 6.941
   Standard error of coefficient of age: 0.135

   Bootstrap with 499:
   Standard error of intercept: 7.9674
   Standard error of coefficient of age: 0.157

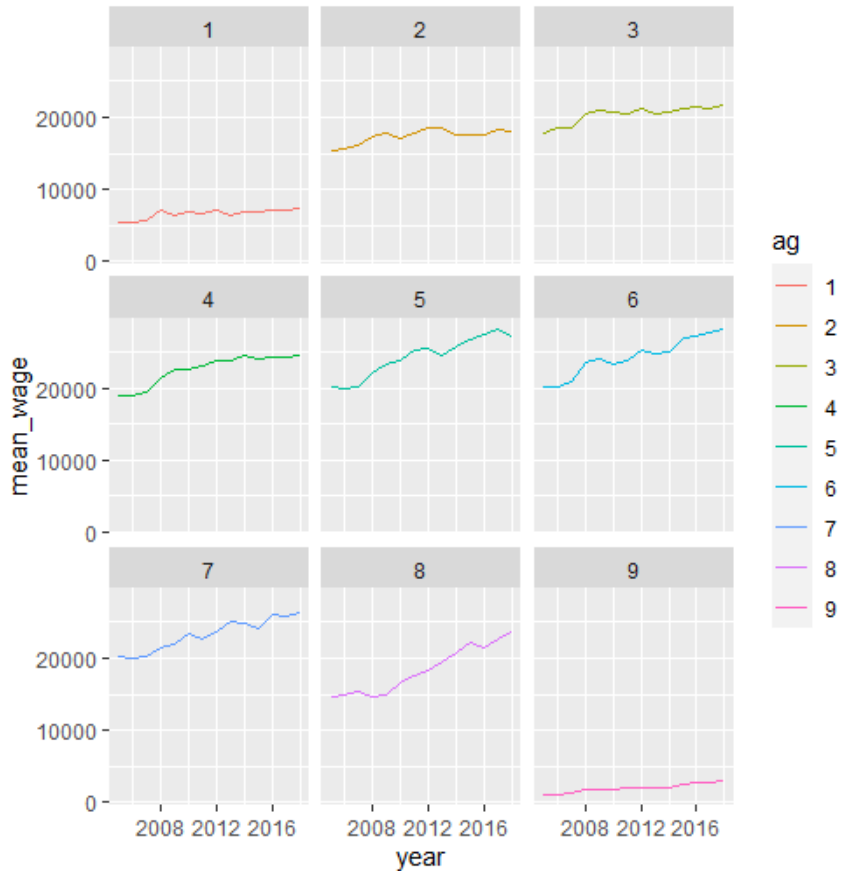
   Bootstrap with 499 replications seems more accurate

Exercise 2 Detrend Data

Q1: Create a categorical variable ag, which bins the age variables into the following groups:

```
> head(ind)
   V1             idind            idmen year  empstat respondent profession gender age  wage ag
1:  1 1120001004058010001 1200010040580100 2005 Inactive          1            Female  31 12334  3
2:  3 1120001006663010001 1200010066630100 2005 Employed          1         38   Male  32 50659  3
3:  4 1120001006663010002 1200010066630100 2005 Employed          0         45 Female  28 19231  2
4:  5 1120001008245010001 1200010082450100 2005  Retired          1            Female  90     0  9
5:  6 1120001008644010001 1200010086440100 2005 Employed          1         34   Male  37 31511  4
6:  7 1120001008644010002 1200010086440100 2005 Employed          0         42 Female  35 24873  3
```

Q2: Plot the wage of each age group across years. Is there a trend?



There is a trend. The mean of wage increases over year.

Q3: After including a time fixed effect, how do the estimated coefficients change?

The estimated coefficient of age changes from -182.4896 to -186.8793

Exercise 3 Numerical Optimization

Q1: Exclude all individuals who are inactive.

```
   V1              idind            idmen year     empstat respondent profession gender age  wage
1:  1 1140001000124010001 1400010001240100 2007 Unemployed          1         NA   Male  49     0
2:  2 1140001000124010002 1400010001240100 2007   Employed          0         52 Female  49 22744
3:  4 1140001001167010001 1400010011670100 2007   Employed          1         21   Male  40  1243
4:  8 1140001002054010001 1400010020540100 2007   Employed          1         22   Male  57     0
5:  9 1140001002054010002 1400010020540100 2007 Unemployed          0         NA Female  54     0
6: 12 1140001005753010001 1400010057530100 2007    Retired          1         NA   Male  71     0
```

Q2: Write a function that returns the likelihood of the probit of being employed.

the loglikelihood is -6582.155

Q3: Optimize the model and interpret the coefficients. You can use pre-programmed optimization packages.

The intercept is 3.8292; the coefficient of age is -0.0679. The coefficient -0.0679 means that an increase of age leads to a decrease in the probability of being employed.

Q4: Can you estimate the same model including wages as a determinant of labor market participation?

Explain.

We can't. The R turns a warning message that algorithm did not converge.

Exercise 4 Discrete choice

Q1: Exclude all individuals who are inactive.

	V1	idind	idmen	year	empstat	respondent	profession	gender	age	wage
1:	3	1120001006663010001	1200010066630100	2005	Employed	1	38	Male	32	50659
2:	4	1120001006663010002	1200010066630100	2005	Employed	0	45	Female	28	19231
3:	5	1120001008245010001	1200010082450100	2005	Retired	1		Female	90	0
4:	6	1120001008644010001	1200010086440100	2005	Employed	1	34	Male	37	31511
5:	7	1120001008644010002	1200010086440100	2005	Employed	0	42	Female	35	24873
6:	8	1120001010299010001	1200010102990100	2005	Employed	1	55	Female	41	30080

Q2: Write and optimize the probit, logit, and the linear probability models.

See the code from line 204 to 269

Q3: Interpret and compare the estimated coefficients. How significant are they?

Probit: intercept: 3.57238; coefficient of age: -0.06359

Logit: intercept: 7.0307; coefficient of age: -0.1241

LPM: intercept: 1.5661; coefficient of age: -0.0201

For probit and logit coefficient, the negative sign means that an increase of age leads to a decrease in the probability of being employed. Their differences are not very large.

For LPM, the -0.0201 means that one unit increase of age leads to 0.0201 decrease of probability of being employed. Unlike probit and logit, the number of LPM coefficient -0.0201 has exact meaning.

For significance：

Probit: t-statistics: -260.078

Logit: t-statistics: -224.173

LPM: t-statistics: -358.72

All three are significant at 1% level.

Exercise 5 Marginal Effects

Q1: Compute the marginal effect of the previous probit and logit models.

Probit: -0.01569

Logit: -0.01224

Q2: Construct the standard errors of the marginal effects.

Probit: 0.000023164698

Logit: 0.00002717307