Econ 613 - Assignment 2

Zixuan Qi

February 5, 2022

Exercise 1

Question 1

The correlation between Y and X is same as the correlation between age and wage because the intercept in matrix X is constant. Thus, the correlation between Y and X is -0.1788512.

Question 2

The coefficient of age is -180.1765, and the intercept is 22075.1066.

Question 3

- When using the standard formulas of OLS, I can get the standard error of intercept is 357.8364 and the standard error of coefficient of age is 6.968824.
- When using bootstrap with 49 and 499 replications respectively, for bootstrap with 49 replications, the standard error of intercept is 358.3699 and the standard error of coefficient of age is 6.980271; for bootstrap with 499 replications, the standard error of intercept is 357.6652 and the standard error of coefficient of age is 6.965084. As the number of replications increases, the results obtained are closer to those obtained by the formulas of the OLS.

Exercise 2

Question 1

The part of the data is as follows. The variable *ag* is objective categorical variable.

	idind	idmen	year	empstat	age	wage	ag
1:	1120001004058010001	1200010040580100	2005	Inactive	31	12334	3
2:	1120001006663010001	1200010066630100	2005	Employed	32	50659	3
3:	1120001006663010002	1200010066630100	2005	Employed	28	19231	2
4:	1120001008245010001	1200010082450100	2005	Retired	90	0	9
5:	1120001008644010001	1200010086440100	2005	Employed	37	31511	4
6:	1120001008644010002	1200010086440100	2005	Employed	35	24873	3

Figure 1: Categorical Variable

Question 2

In this figure, group 1 is the age from 18 to 25; group 2 is is the age from 26 to 30; group 3 is the age from 31 to 35; group 4 is is the age from 36 to 40; group 5 is the age from 41 to 45; group 6 is is the age from 46 to 50; group 7 is the age from 51 to 55; group 8 is is the age from 56 to 60; group 9 is is the age larger than 60.

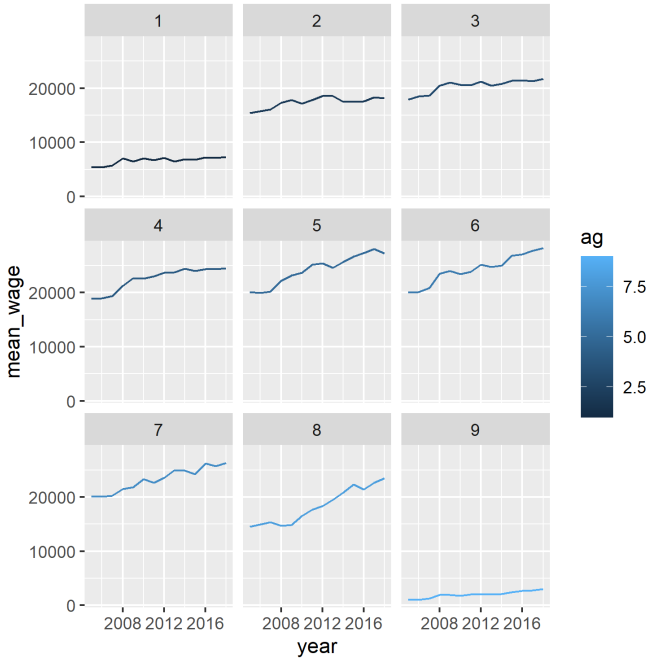

Figure 2: Wage of Each Age Group

As you can see in the figure, the mean of wage is increasing over year.

Question 3

After including a time fixed effect, the coefficient of age changes from -182.4896 to -186.8793.

Exercise 3

Question 1

The part of data that exclude all individuals who are inactive is as follows.

	idind	idmen	year	empstat	age	wage
1:	11400010000124010001	1400010001240100	2007	Unemployed	49	0
2:	11400010000124010002	1400010001240100	2007	Employed	49	22744
3:	11400010001167010001	1400010011670100	2007	Employed	40	1243
4:	11400010002054010001	1400010020540100	2007	Employed	57	0
5:	11400010002054010002	1400010020540100	2007	Unemployed	54	0
6:	11400010005753010001	1400010057530100	2007	Retired	71	0

Figure 3: No Inactive

Question 2

The log likelihood is equal to -6558.221.

Question 3

The coefficient of age is -0.06820941, and the intercept is 3.84885488. The curve of the probit function is S-shaped curve, and the coefficients control the position and shape of the curve. The corresponding slopes of age at different positions on the curve are also different. And when other variables maintain same but age increases, the probability of being employed decrease.

Question 4

No. There is warning message showing that algorithm does not converge. I think the reason is that including wages as a determinant of labor market participation leads to an over-fitting problem: too many variables in the model. The consequences is that model likelihood is not defined, and thus I can't get the model to converge.

Exercise 4

Question 1

The part of data from 2005 to 2015 that exclude all individuals who are inactive is as follows.

	idind	idmen	year	empstat	age
1:	1120001006663010001	1200010066630100	2005	Employed	32
2:	1120001006663010002	1200010066630100	2005	Employed	28
3:	1120001008245010001	1200010082450100	2005	Retired	90
4:	1120001008644010001	1200010086440100	2005	Employed	37
5:	1120001008644010002	1200010086440100	2005	Employed	35
6:	1120001010299010001	1200010102990100	2005	Employed	41

Figure 4: No Inactive

Question 2

The function and the optimization approach can be seen in the code. In the probit model, the coefficient of age is -0.06358964, and the intercept is 3.57237881. In the logit model, the coefficient of age is -0.1241425, and the intercept is 7.0307021. In the Linear Probability model, the coefficient of age is -0.01846614, and the intercept is 1.54460691.

Question 3

In the question 2, I have already reported the coefficients of three ways. As you can see, the coefficients of age of the three models are not very different, and the signs of the coefficients are same. In probit and logit model, since the relationship between age and the probability of being employed is nonlinear, the curve of the logit and probit functions are S-shaped curve, and the estimators control the position and shape of the curve. The corresponding slopes of age at different positions on the curve are also different. And when other variables maintain same but age increases, the probability of being employed decrease. In LPM, the coefficient of age means that one unit increase in age results in 0.01846614 unit decrease in the probability of being employed.

In the probit model,the t-statistics of the coefficient of age is -260.0777. In the logit model,the t-statistics of the coefficient of age is -224.1728. In the LPM, the t-statistics of the coefficient of age is -381.5322. Thus, these three coefficients are significant at 1% level.

Exercise 5

Question 1

The marginal effect of the previous probit model is -0.01568521 and the marginal effect of the previous logit model is -0.01600454.

Question 2

In the previous probit model, the standard error of marginal effect is 0.0000258. In the previous logit model, the standard error of marginal effect is 0.0000379.