

# Econ 613 A2

## Peilin Wang

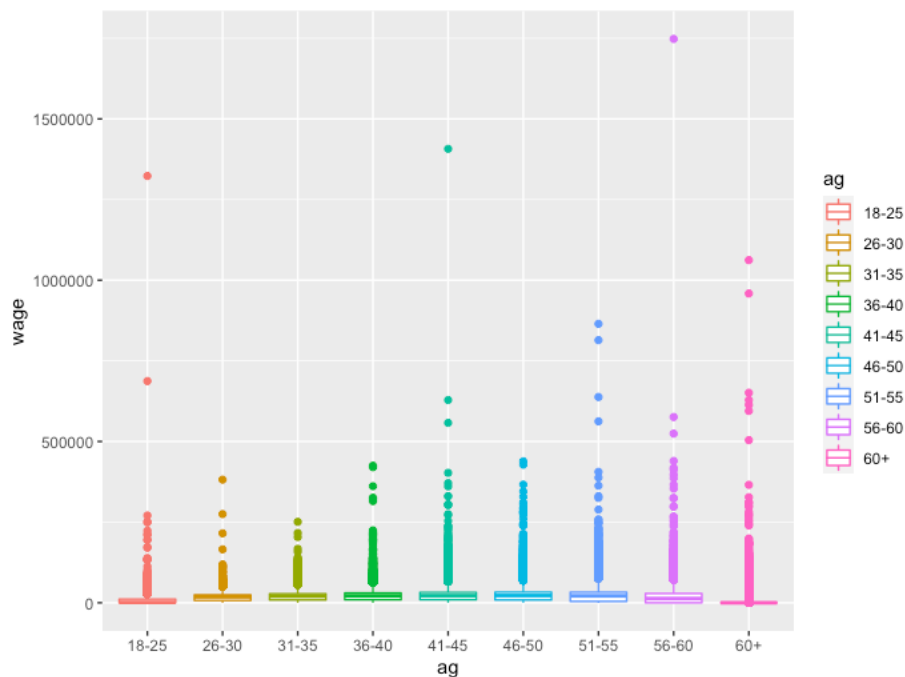
### Exercise 1:

- (1) Calculate the correlation between Y and X: 0.1579991
- (2) Calculate the coefficients on this regression: 568.5085
- (3) Calculate the standard errors of  $\beta$ :
  - OLS: 1.432955
  - bootstrap with 49: 1.618131
  - bootstrap with 499: 1.654581

The method that we use the standard formula of OLS to calculate standard error relies on the normality assumption of error term. If the residuals are surly normally distributed, the answer will be accurate. The bootstrap method is useful if we don't trust the distribution that we've assumed. For this question, we can see that the answers are close but not the same. So, the error term is close to normal distribution.

### Exercise 2:

- (1) Create a categorical variable ag: please check in R
- (2) Plot the wage of each age group across years. Is there a trend?



Yes, there is a trend. The wage will increase when age increases. However, when people are getting old, wage of some people will decrease, and others will remain high.

(3) After including a time fixed effect, how do the estimated coefficients change?

```

              [,1]
10625.07352
 297.81653
y_6      114.58388
y_7      135.20977
y_8      -90.79323
y_9       758.23827
y_10     568.06612
y_11    1106.64241
y_12    1550.65094
y_13    1424.06266
y_14    2029.96350
y_15    2305.74479
y_16    2982.29116
y_17    2940.21715
y_18    3044.37546
> |

```

The coefficient of age decreases when adding year dummy variables.

### Exercise 3

(1), (2): please check in R

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

```

> out1[which(out1$V3 == min(out1$V3)),
      V1      V2      V3
81 0.7235125 0.013188 49853.31
> |

```

Older people are more likely to be employed comparing to reference group with everything else being equal.

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

	V1	V2	V3	V4
85	-2.94056	0.05079909	0.0002768177	71851.26

No, we cannot estimate the same model including wages as a determinant. Since there are some unemployed people who don't have wages and others who have high wages, the outliers may exist. So, the wage should be removed.

#### Exercise 4

(1) : please check in R

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

- Probit:

```
> print(beta_prob)
      V1      V2      V3      V4      V5      V6      V7      V8
53 0.7628026 0.01196652 0.0171697 0.08081854 0.1100723 0.0269495 0.02204228 0.05555788
      V9      V10      V11      V12      V13
53 0.01066392 -0.03945226 -0.03275415 8.671153 56622.5
> |
```

- Logit:

	V1	V2	V3	V4	V5	V6	V7	V8
58	1.12029	0.02530116	0.03240617	0.1573336	0.2133244	0.04619446	0.03798792	0.1013241
	V9	V10	V11	V12	V13			
58	0.01191332	-0.0838495	-0.07076045	-0.1113917	42213.76			

- Linear probability:

```

                                [,1]
                                0.7977483650
                                0.0023358617
y.6    0.0029332879
y.7    0.0139479286
y.8    0.0184425587
y.9    0.0040834132
y.10   0.0033035718
y.11   0.0088873932
y.12   0.0008988494
y.13  -0.0083476676
y.14  -0.0070497876
y.15  -0.0109176395
> |

```

(3) Interpret and compare the estimated coefficients. How significant are they?

Probit SD:

```

[1] 0.022685524 0.000405033 0.022776655 0.022933998 0.023153899 0.022685561 0.022486490
[8] 0.022543295 0.022023540 0.022265421 0.022252321 0.022239508
> |

```

Logit SD:

```

> print(sigma_logit)
[1] 0.0438462273 0.0008092424 0.0440143386 0.0447322972 0.0453176586 0.0438509935
[7] 0.0434687814 0.0437630631 0.0425098713 0.0427200106 0.0427735798 0.0426500171
> |

```

Linear SD:

```

                                y.6      y.7      y.8      y.9
0.00418417347 0.00007411822 0.00408558482 0.00404926759 0.00405670764 0.00405805470
      y.10      y.11      y.12      y.13      y.14      y.15
0.00402439205 0.00400028914 0.00394855509 0.00403494842 0.00402191226 0.00403490086
> |

```

By checking the standard error, we can see that the variables have small SE which means that the sample mean accurately reflects the population mean. Even though the coefficients are different in these three methods, we can still think that the results are significant. I think this question also needs to report significant level, but I cannot figure it out.

## Exercise 5

(1) Compute the marginal effect of the previous probit and logit models.

- Probit:

```
53
V1    0.122464072
V2    0.001921164
V3    0.002756507
V4    0.012975004
V5    0.017671549
V6    0.004326605
V7    0.003538776
V8    0.008919534
V9    0.001712038
V10   -0.006333860
V11   -0.005258511
V12   1.392109411
> |
```

- Logit:

```
58
V1    0.0469221056
V2    0.0010597106
V3    0.0013572960
V4    0.0065897416
V5    0.0089348522
V6    0.0019348029
V7    0.0015910810
V8    0.0042438478
V9    0.0004989759
V10   -0.0035119421
V11   -0.0029637216
V12   -0.0046655148
> |
```

(2) Construct the standard errors of the marginal effects.