# ECON613 HW4

Yuanhang Zhao

April 2022

## 1 Exercise 1

### 1.1 Visualization

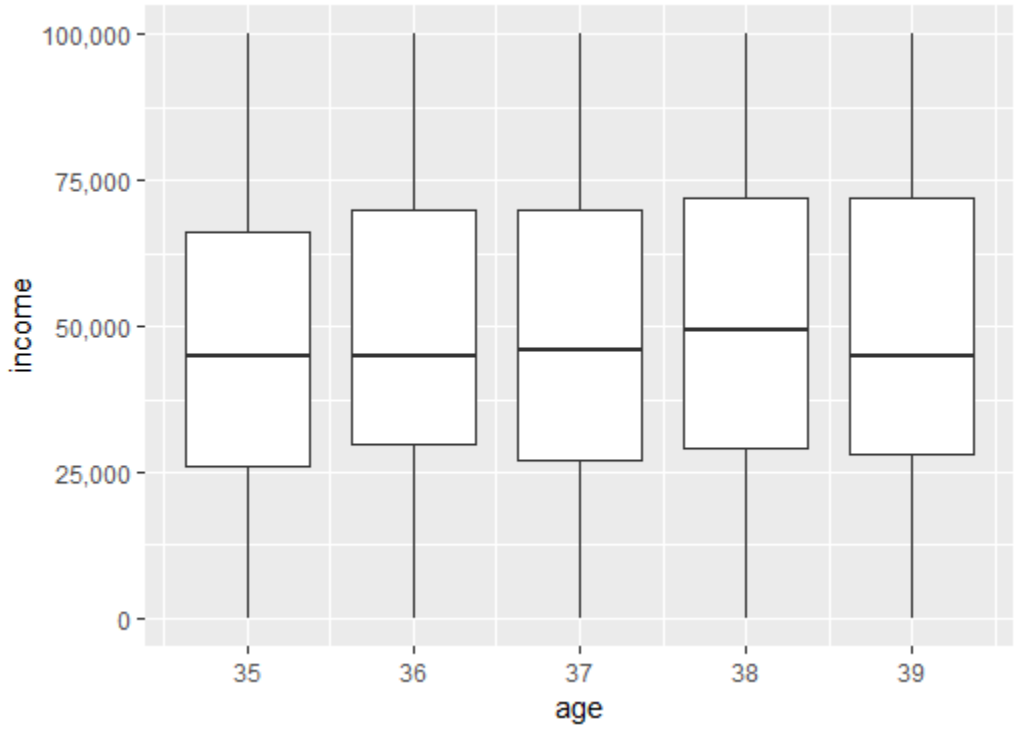


Figure 1: Distribution group by age

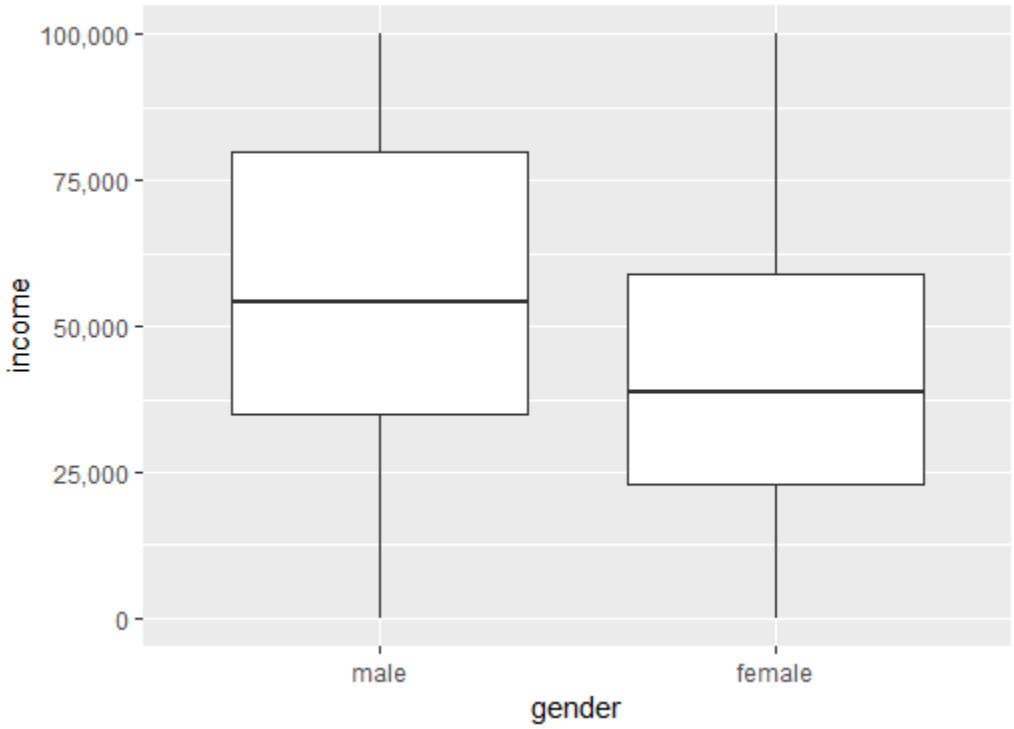


Figure 2: Distribution group by gender

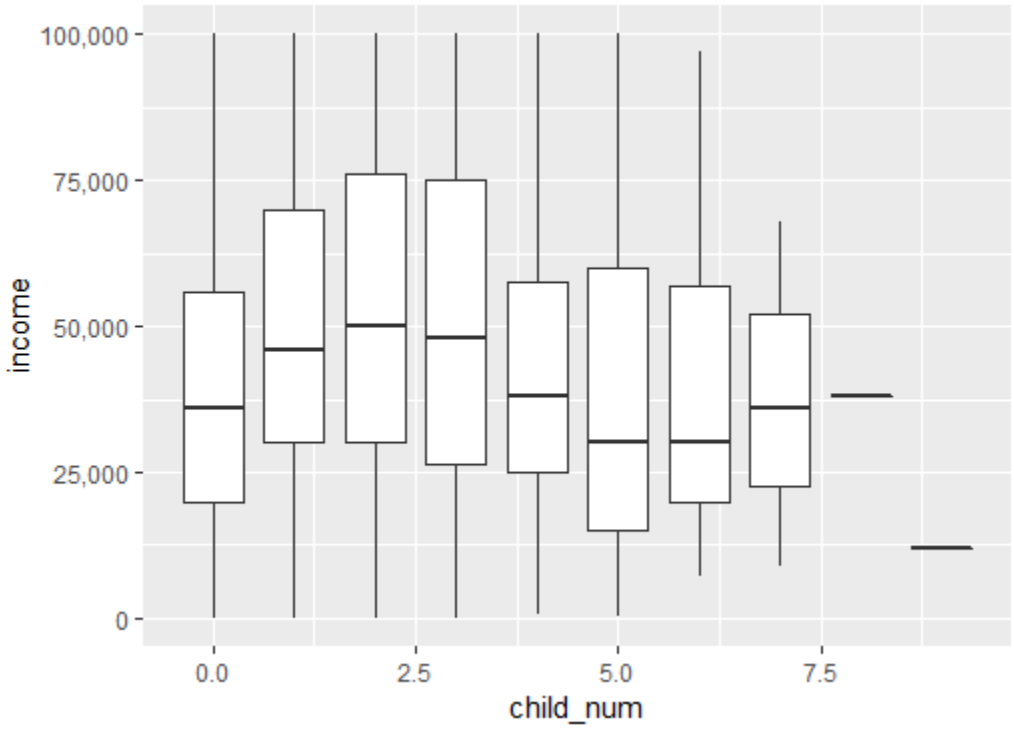


Figure 3: Distribution group by number of child

Table 1: group by age

| | age | income_zero |
|---|---|---|
| 1 | 35.00 | 0.40 |
| 2 | 36.00 | 0.39 |
| 3 | 37.00 | 0.40 |
| 4 | 38.00 | 0.41 |
| 5 | 39.00 | 0.41 |

Table 2: group by gender

| | gender | income_zero |
|---|---|---|
| 1 | male | 0.40 |
| 2 | female | 0.41 |

Table 3: group by marstat and number of child

| | marital_status | child_num | income_zero |
|---|---|---|---|
| 1 | never | 0 | 0.42 |
| 2 | never | 1 | 0.26 |
| 3 | never | 2 | 0.27 |
| 4 | never | 3 | 0.32 |
| 5 | never | 4 | 0.44 |
| 6 | never | 5 | 0.37 |
| 7 | never | 6 | 0.60 |
| 8 | never | 7 | 1.00 |
| 9 | never | 9 | 0.00 |
| 10 | never | | 0.23 |
| 11 | married | 0 | 0.25 |
| 12 | married | 1 | 0.14 |
| 13 | married | 2 | 0.17 |
| 14 | married | 3 | 0.20 |
| 15 | married | 4 | 0.26 |
| 16 | married | 5 | 0.36 |
| 17 | married | 6 | 0.25 |
| 18 | married | 7 | 0.25 |
| 19 | married | 8 | 0.50 |
| 20 | married | | 0.11 |
| 21 | seperated | 0 | 0.47 |
| 22 | seperated | 1 | 0.35 |
| 23 | seperated | 2 | 0.25 |
| 24 | seperated | 3 | 0.33 |
| 25 | seperated | 4 | 0.38 |
| 26 | seperated | 5 | 0.50 |
| 27 | seperated | 6 | 0.00 |
| 28 | seperated | | 0.22 |
| 29 | divorced | 0 | 0.29 |
| 30 | divorced | 1 | 0.17 |
| 31 | divorced | 2 | 0.20 |
| 32 | divorced | 3 | 0.18 |
| 33 | divorced | 4 | 0.28 |
| 34 | divorced | 5 | 0.33 |
| 35 | divorced | | 0.20 |
| 36 | widowed | 0 | 0.00 |
| 37 | widowed | 1 | 0.29 |
| 38 | widowed | 2 | 0.25 |
| 39 | widowed | 5 | 0.00 |
| 40 | widowed | | 0.29 |
| 41 | | 0 | 0.50 |
| 42 | | 1 | 0.55 |
| 43 | | 2 | 0.50 |
| 44 | | 3 | 0.75 |
| 45 | | 4 | 1.00 |
| 46 | | 6 | 1.00 |
| 47 | | | 1.00 |

Table 4: group by marstat

| | marital_status | income_zero |
|---|---|---|
| 1 | never | 0.28 |
| 2 | married | 0.17 |
| 3 | seperated | 0.34 |
| 4 | divorced | 0.22 |
| 5 | widowed | 0.24 |
| 6 | | 0.99 |

Table 5: group by number of child

| | child_num | income_zero |
|---|---|---|
| 1 | 0 | 0.36 |
| 2 | 1 | 0.19 |
| 3 | 2 | 0.20 |
| 4 | 3 | 0.23 |
| 5 | 4 | 0.31 |
| 6 | 5 | 0.36 |
| 7 | 6 | 0.42 |
| 8 | 7 | 0.40 |
| 9 | 8 | 0.50 |
| 10 | 9 | 0.00 |
| 11 | | 0.62 |

### 1.2 Interpretation

1. Elder people have a little higher wage in average(not significant).
2. Men have higher income than women.
3. The distribution on number of child is an anti-smile curve, which means people who have around 2 or 3 kids will have higher wage.
4. Zero values cannot be ignored(zeros have large share of the income data).

## 2 Exercise 2

### 2.1 OLS Model

Table 6: OLS result

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 5273.3452 | 14284.7723 | 0.37 | 0.7120 |
| age | 126.2670 | 382.5048 | 0.33 | 0.7414 |
| genderfemale | -19926.5023 | 1083.3399 | -18.39 | 0.0000 |
| marriedTRUE | 4170.5914 | 1200.8733 | 3.47 | 0.0005 |
| child_num | 434.6118 | 508.2162 | 0.86 | 0.3926 |
| edu | 2165.1072 | 157.7276 | 13.73 | 0.0000 |
| res_edu | 492.8864 | 98.9243 | 4.98 | 0.0000 |
| work_exp | 1082.4174 | 98.6077 | 10.98 | 0.0000 |

### Interpretation

1. When controlling other variables, women have lower wage than men.
2. When controlling other variables, people have longer education year earn more.
3. When controlling other variables, people whose parent have higher education level have higher wage, but this effect is smaller than their own education level.
4. When controlling other variables, married people generally have higher wage.
5. When controlling other variables, people who have more work experience have higher wage.
6. The effects of age and number of kid are not significant.

### Selection Problem in OLS

There are a "cannot be ignored" share of zero and missing value in income variable, which is potentially represent the potential income of those individuals. If we just use OLS to estimate, selection problem occurred and the results may be biased.

### 2.2 Heckman Selection - First stage

Table 7: First stage result and comparison

| | heckit : est | heckit :se | without package : est | without package :se |
|---|---|---|---|---|
| (Intercept) | 1.36 | 2.45 | 1.36 | 2.41 |
| age | 0.03 | 0.06 | 0.03 | 0.06 |
| genderfemale | -0.32 | 0.19 | -0.32 | 0.19 |
| marriedTRUE | -0.47 | 0.23 | -0.47 | 0.23 |
| child_num | 0.08 | 0.09 | 0.08 | 0.09 |
| edu | 0.03 | 0.02 | 0.03 | 0.02 |
| res_edu | -0.02 | 0.02 | -0.02 | 0.02 |
| work_exp | 0.03 | 0.02 | 0.03 | 0.02 |

### 2.3 Heckman Selection - Second stage

Table 8: Second stage comparison

| | heckit : est | heckit :se | without package : est | without package :se |
|---|---|---|---|---|
| (Intercept) | 23987.01 | 31306.85 | 23988.26 | 31050.89 |
| age | -255.25 | 834.71 | -255.27 | 833.22 |
| genderfemale | -16123.76 | 2311.09 | -16123.37 | 2234.40 |
| marriedTRUE | 9315.48 | 2494.68 | 9315.96 | 2392.62 |
| child_num | -499.64 | 1124.01 | -499.72 | 1106.37 |
| edu | 1734.98 | 335.55 | 1734.96 | 333.66 |
| res_edu | 687.08 | 215.25 | 687.09 | 215.04 |
| work_exp | 791.34 | 203.48 | 791.31 | 208.19 |

#### Interpretation

Compared with OLS model, the sign are similar in Heckman result, so the interpretation are similar. However, the scale of the coefficents changes, which is caused by considering the selection problem.

## 3 Exercise 3

### 3.1 Censored value

With the figure shows the distribution of the income value, we can find that the censored value is supposed to be 100,000.

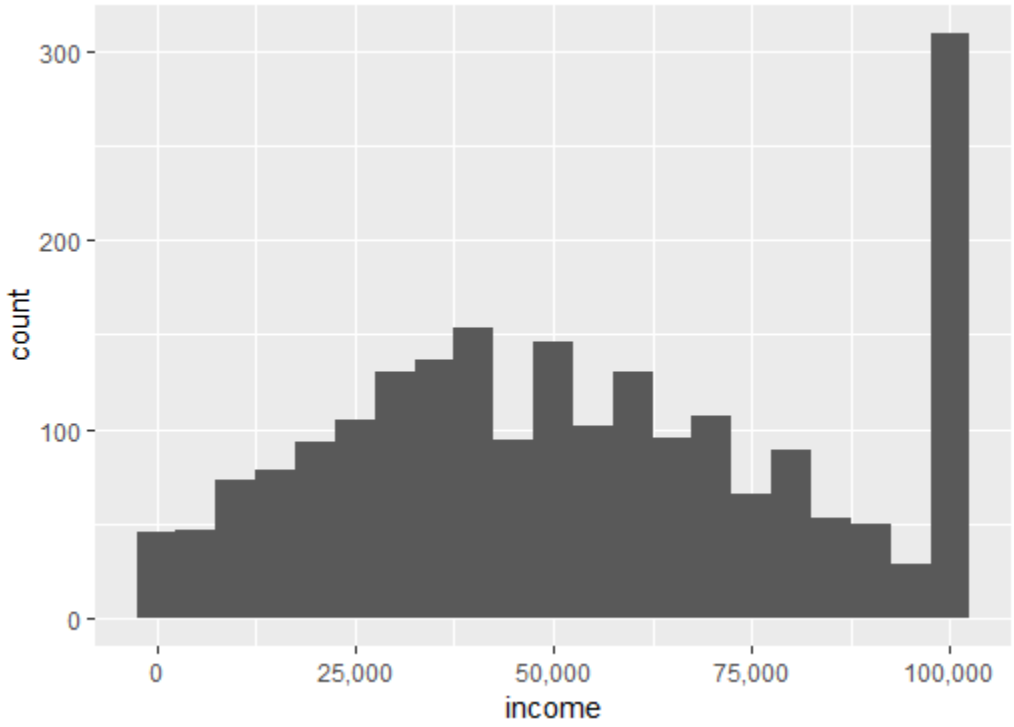


Figure 4: Distribution of income

### 3.2 Censored Problem

We can solve the censored model with a tobit model, which is consider the censored value as zero. In this data, the equation is as followed:

$$income = \begin{cases} 0, & income = 100,000, \\ income, & income < 100,000. \end{cases} \tag{1}$$

With the result from tobit model, we can find that the sign of coefficient are similar to the results from exercise 2. The changes in scale of estimates results from the change in censored value.

Table 9: Results from tobit model

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -2472.60 | 16198.64 | -0.15 | 0.88 |
| age | 234.95 | 435.95 | 0.54 | 0.59 |
| genderfemale | -22424.83 | 1236.83 | -18.13 | 0.00 |
| marriedTRUE | 4694.43 | 1350.23 | 3.48 | 0.00 |
| child_num | 587.50 | 586.15 | 1.00 | 0.32 |
| edu | 2385.20 | 187.06 | 12.75 | 0.00 |
| res_edu | 595.53 | 112.92 | 5.27 | 0.00 |
| work_exp | 1181.21 | 122.90 | 9.61 | 0.00 |
| Log(scale) | 10.23 | 0.02 | 569.63 | 0.00 |

## 4 Exercise 4

Individuals have their own abilities, which cannot be observed but correlated with the income. When regress their wage with the determinants, the estimation is biased. Moreover, the panel data is the data over time, in every period, the unobserved error are not independent, which is also the bias.

### 4.1 Within

Table 10: Within Group Effect

| | Estimate | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| age | 2058.17 | 21.79 | 94.44 | 0.00 |
| marriedothers | -5908.93 | 247.56 | -23.87 | 0.00 |
| work_exp | 916.00 | 32.86 | 27.88 | 0.00 |
| edu | 3517.28 | 70.27 | 50.05 | 0.00 |

### 4.2 Between

Table 11: Between Group Effect

| | Estimate | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -57300.94 | 2390.11 | -23.97 | 0.00 |
| age | 936.00 | 63.44 | 14.75 | 0.00 |
| gendermale | 9217.06 | 362.39 | 25.43 | 0.00 |
| marriedothers | -7976.42 | 574.19 | -13.89 | 0.00 |
| work_exp | 1574.16 | 76.61 | 20.55 | 0.00 |
| edu | 4159.41 | 124.96 | 33.29 | 0.00 |

### 4.3 First Difference

Table 12: First difference strategy

| | Estimate | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 1214.59 | 113.65 | 10.69 | 0.00 |
| age | 1829.99 | 54.06 | 33.85 | 0.00 |
| marriedothers | -1567.38 | 246.40 | -6.36 | 0.00 |
| work_exp | 677.71 | 33.03 | 20.52 | 0.00 |
| edu | 872.54 | 87.34 | 9.99 | 0.00 |

## Interpretation

The interpretation for the results above are similar, because they have same signs in coefficients. The difference comes from the difference in groups. Within group result shows the effects of the determinants on individual level. Between group shows the effects between different individuals. First different can control the individual heterogeneity.