HW4

Rulin Wang

Exercise1

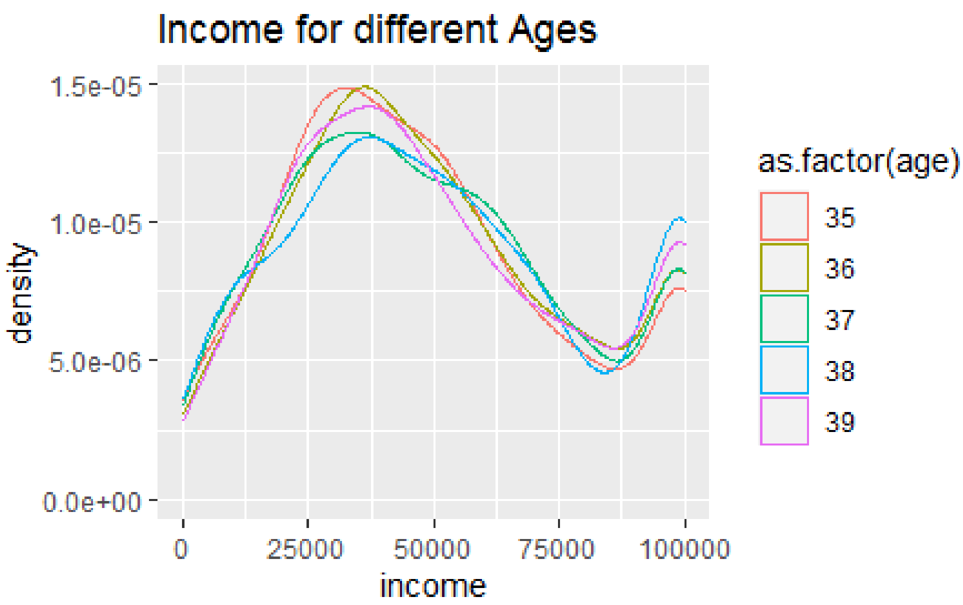
(1) data1$work_exp

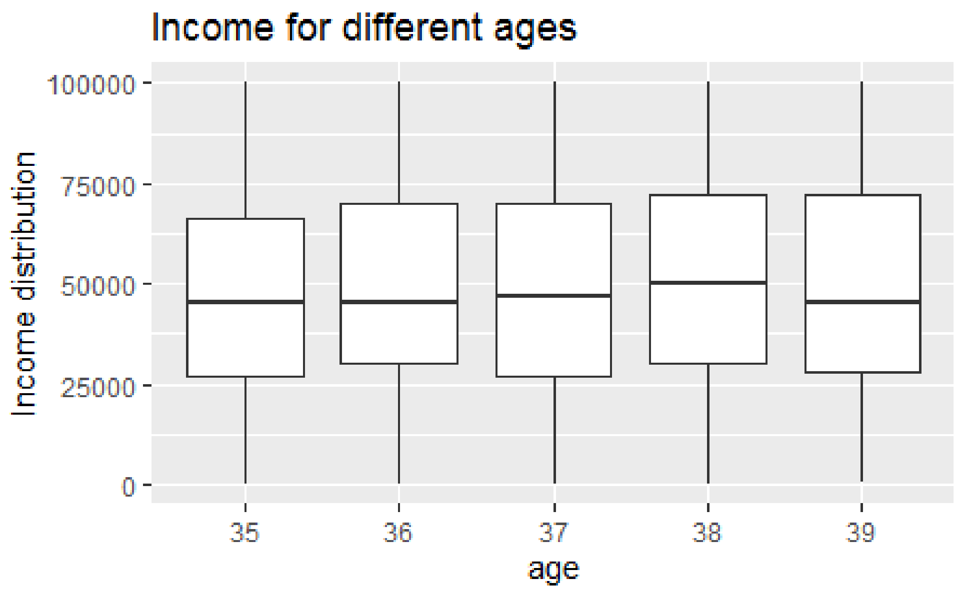
	age	work_exp
	38	0.0000000
	37	12.4230769
	36	1.6923077
	38	1.9230769
	37	13.4615385
	37	2.2500000
	36	2.3653846
	38	4.1923077
	37	3.2307692
	35	5.0769231
	37	11.9423077
	38	14.9230769
	35	0.0000000

(2) "edu_year" in "data1" is years of schooling of each individual

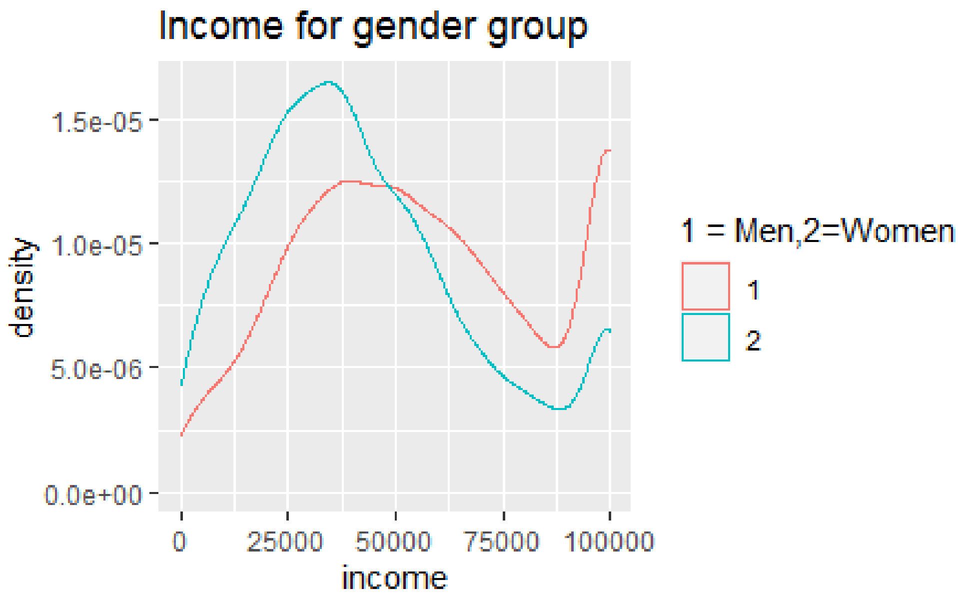
edu_year
NA
12
16
12
12
12
0
16
18
18
16
12
12

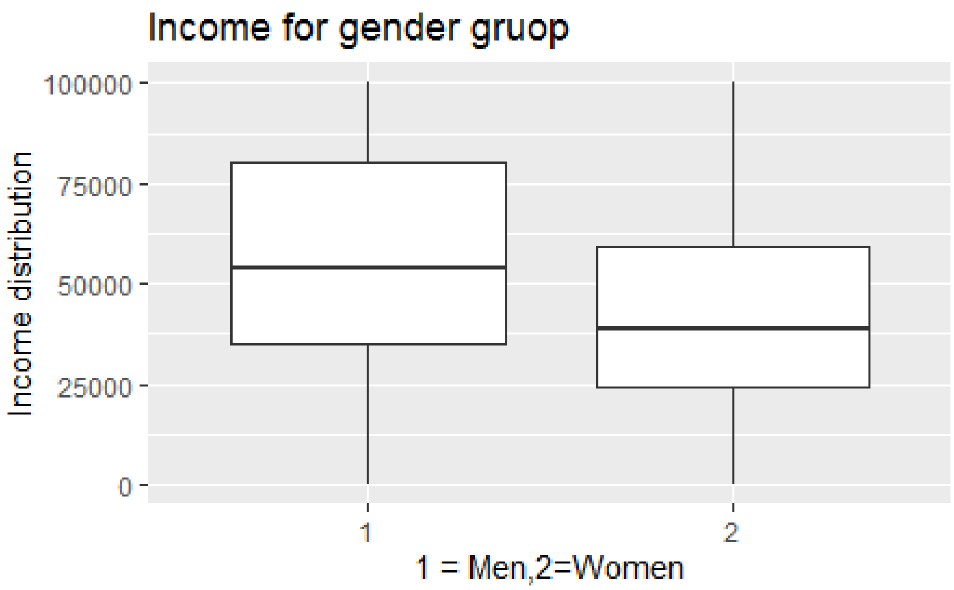
(3) I) income by age



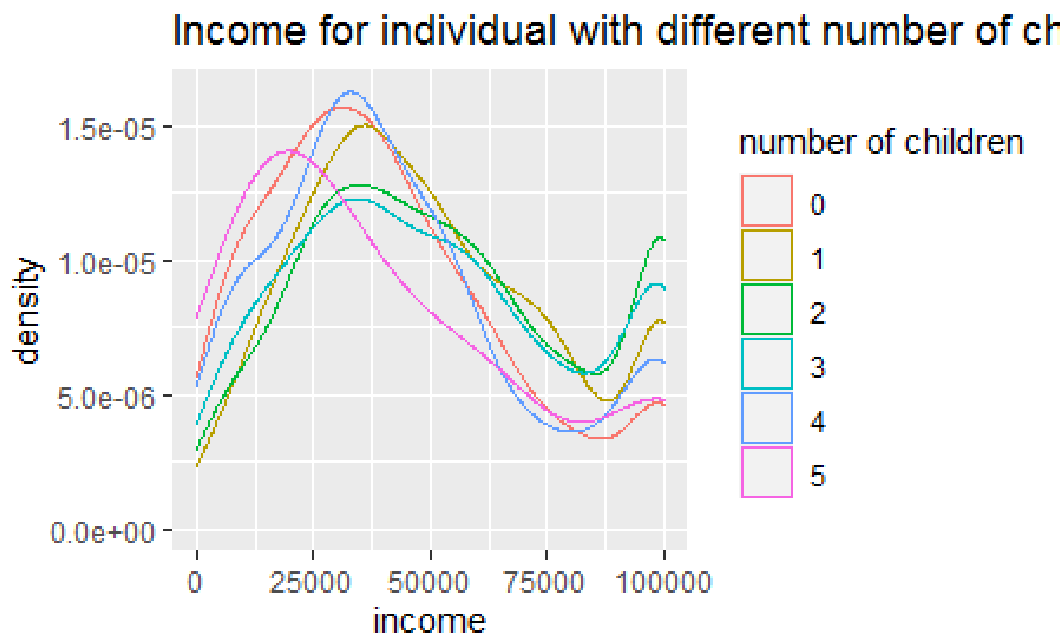


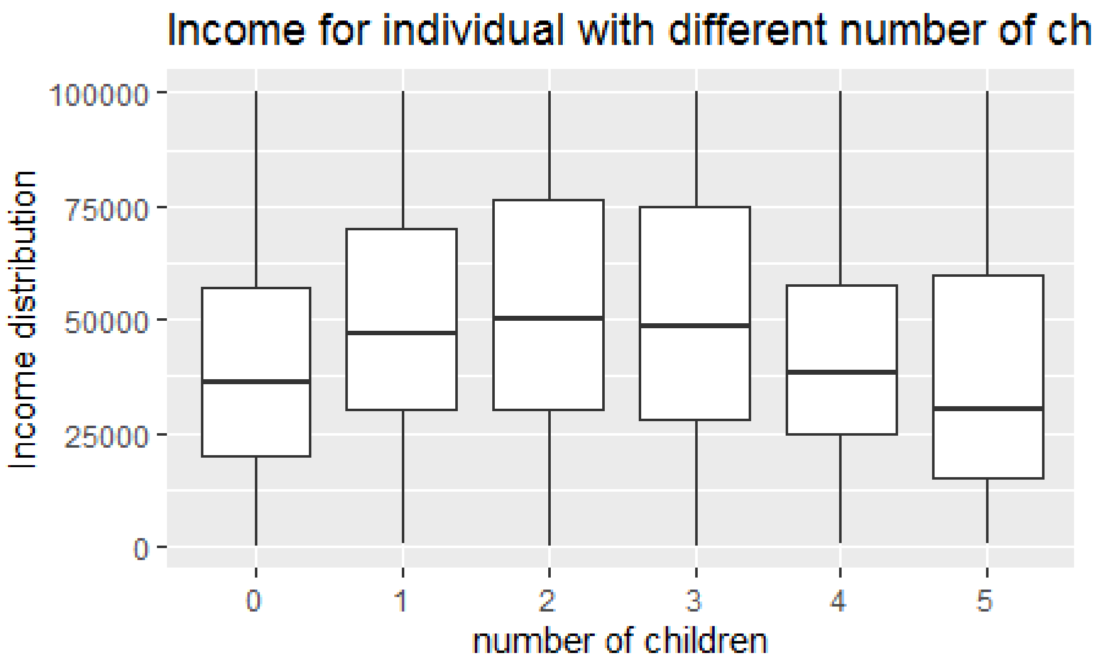
ii) income by gender





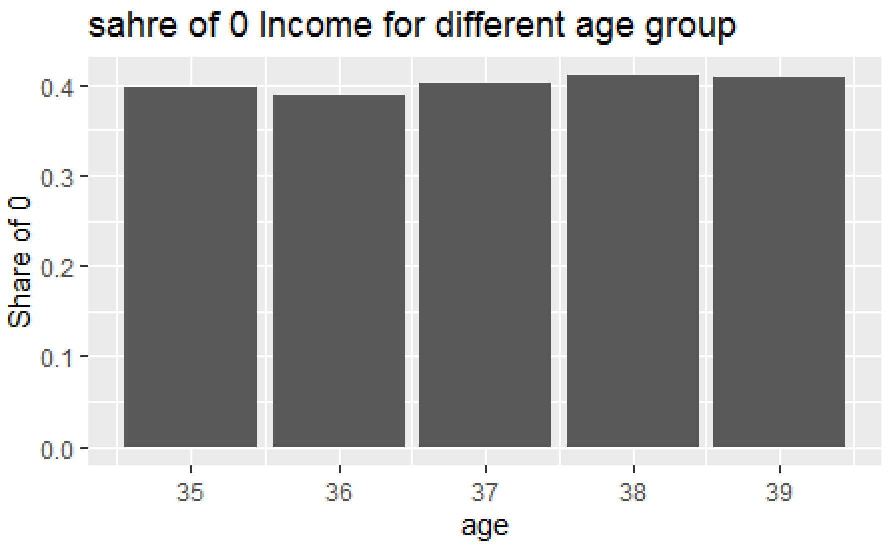
iii) income by number of children



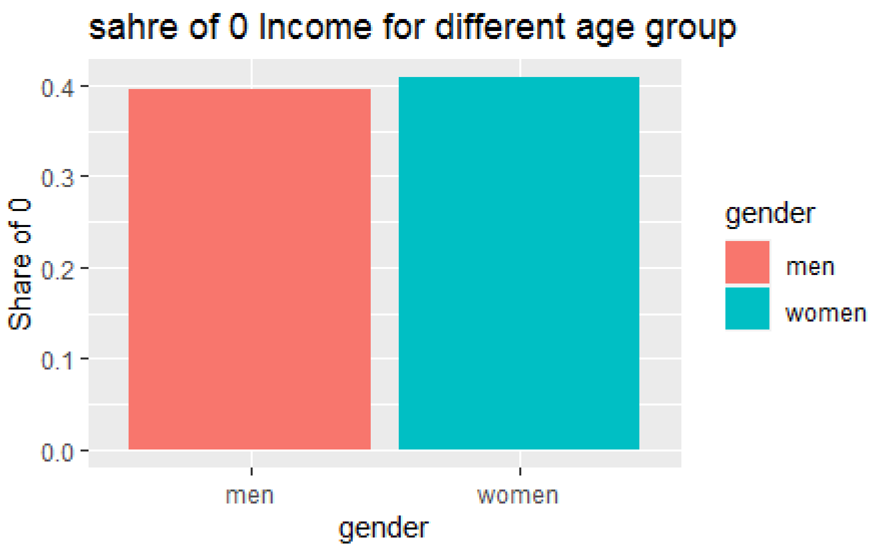


The share of 0 income

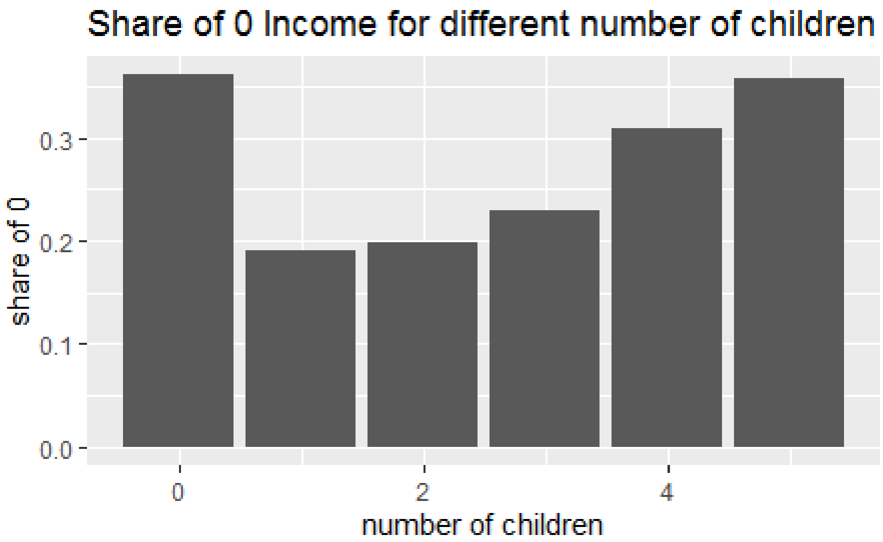
(I)Income by age



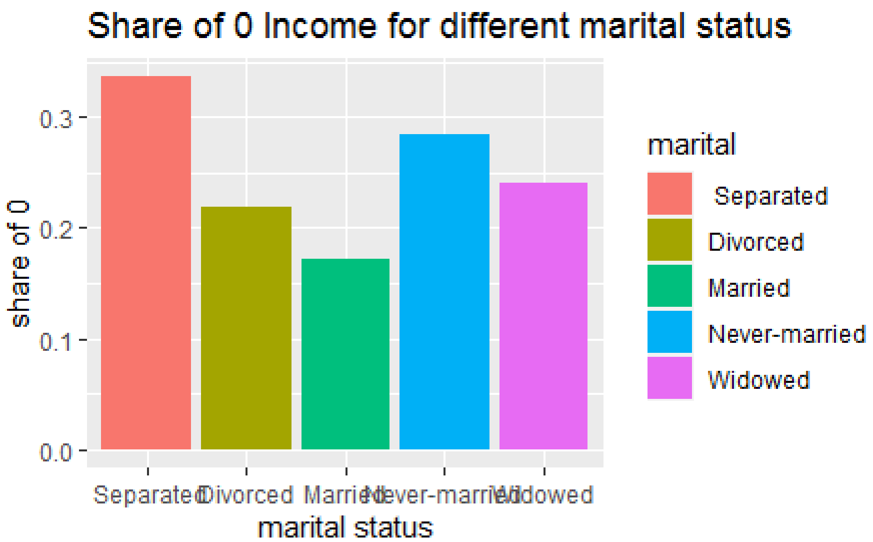
ii)Income by gender



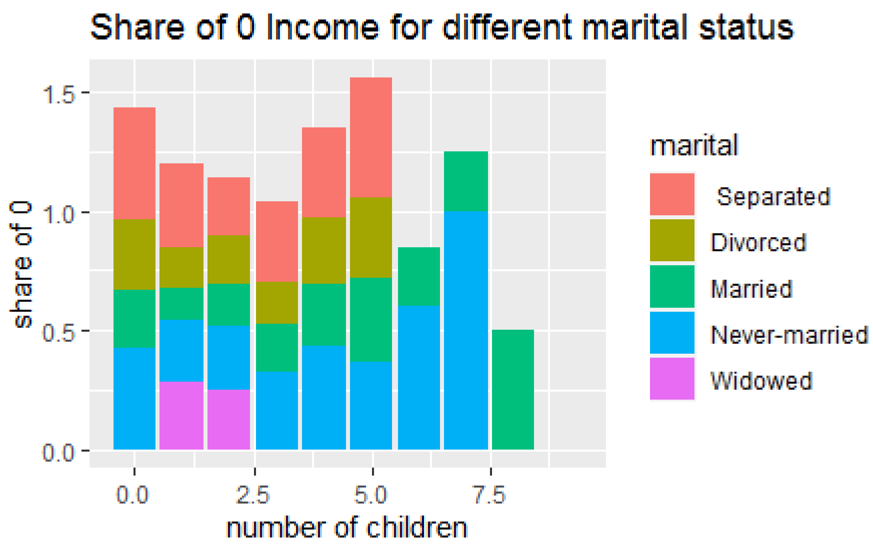
iii)Income by number of children



IV) Income by marital status



combined children and marital status



Interpretation

(1) For respondents aged 35 to 39, there was little difference in income distribution. 38-year-old respondents earned slightly more than the other groups;

#(2) For respondents, men generally earn more than women;

#(3) Respondents with 1-3 children had higher incomes than those with more than 3 children and no children;

#(4) For respondents aged 35 to 39, their 0 income ratio was similar, at about 0.4;

#(5)For respondents, men have the similar 0 income ratio as women, at about 0.4;

#(6) Respondents with no children and 4-5 children have a higher 0 income ratio than those with 1-3 children
Seperated and never-married respondents have a higher 0 income ratio than others.
Never-married respondents with 6-7 children have a very high 0 income ratio.

Exercise2

(1)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11124.13    1290.65   8.619   <2e-16 ***
work_exp     1068.32      66.10  16.163   <2e-16 ***
edu_year     2341.17      88.53  26.443   <2e-16 ***
```

Working experience and year of education both have postive effect on income.

For each additional year of work experience, earnings rise by about 1068$

For each additional year of education, earnings rise by about 2341$

But there might be a selection problem in this model because people who report 0 income or don't report may not be random.

For example, people with high income may not be willing to report their income.

Thus income data may have error.

(2) Heckman model can deal with this selection problem because in the first step, Heckman used a probit model to calculate the probability of reporting income. In the second step, Heckman combined these predicted individual probabilities into an additional explanatory variable, along with other control variables, to solve the selection problem.

(3)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   69028.3     4434.2  15.567  < 2e-16 ***
work_exp       -687.6      171.0  -4.021 5.86e-05 ***
edu_year        241.3      143.2   1.685   0.0921 .  
imr          -75478.8     4605.5 -16.389  < 2e-16 ***
---
```

imr is significant. The results of Heckman are quite different from those of OLS.

Many low-income people may report 0 income or not report income, which leads to overly optimistic OLS results.

Exercise3

(1) ## 637 people have 100,000 income . I see 0 and 100,000 repeated many times. Repeating 0 is reasonable because many people don't have income. But 100,000 is unreasonable.

I check there are notes showing that reports used truncated values. So it might be 100,000.

(2) Tobit model can be used to solve this problem. Independent variables: income dependent variables: work year; education year

(3)

```
> res_2$par
[1]  1861.64494   2259.66374   2387.45786    10.34778 -3079.48590   2024.80230   2382.51682
>
```

(4) Interpretation: Income has a positive correlation with work experience and education year. Compared to the original model, coefficients are larger. This is because when using censor data.

We ignore some large data, then we underestimate the coefficients.

Exercise4

(1) Education, work experience and marital status are determinants of wages. But ability may create a selection bias. Because people with higher ability can get higher grade then study longer.

Also, people with higher ability may perform better in marriage market, and they can work more years due to their ability.

And people with higher ability usually have a higher salary.

Thus ability is related to both dependent and independent variables, resulting in endogenous problems.

(2) (i)between estimator:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -19371.21    1316.08  -14.72   <2e-16 ***
edu           2347.94     104.50   22.47   <2e-16 ***
exper         2154.46      67.46   31.94   <2e-16 ***
ms            3211.77     237.83   13.51   <2e-16 ***
---
```

(ii) Within estimator

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.599e-12  4.934e+01    0.00        1    
data6$edu1   4.951e+03  4.667e+01  106.09   <2e-16 ***
data6$exp1   2.332e+03  3.105e+01   75.10   <2e-16 ***
data6$ms1    7.234e+03  1.286e+02   56.26   <2e-16 ***
```

(iii) Difference estimator

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3210.16      56.79   56.52   <2e-16 ***
data8$edu2   1162.83      64.16   18.12   <2e-16 ***
data8$exp2    741.67      34.41   21.55   <2e-16 ***
data8$ms2    1554.63     159.29    9.76   <2e-16 ***
```

(3) The coefficients obtained by three models are different. But all coefficients are significant and positive. Thus we can make sure education, work experience and marital status have positive effects on wages. In addition, I believe that the three models produce different coefficients because of the different ways in which they deal with the data. Different approaches to the problem of ability result in different results.