Assignment Four Outputs

Exercise 1

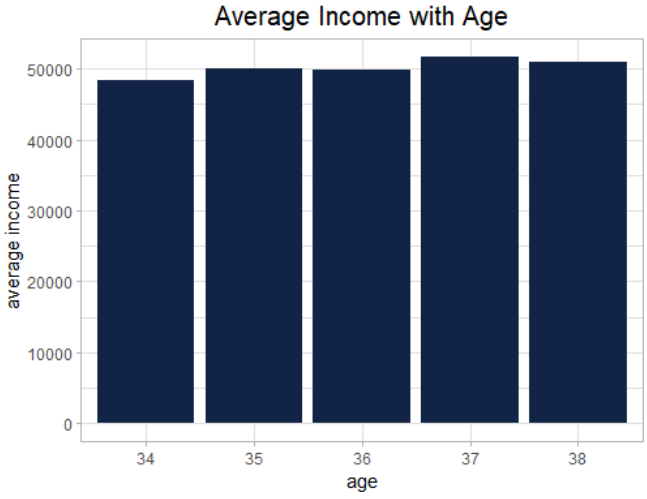
1. For the *age* variable, I use the formula: 2019 -"KEY_BDATE_Y_1997"-1 to make sure everyone is from 12 to 16 years old in 1997 (as described in the file). For the *work_exp* variable, I add all working weeks and then divide it by 52. Here provides the head of these two variables.

age	work_exp
37	0.0000000
36	12.4230769
35	1.6923077
37	1.9230769
36	13.4615385
36	2.2500000

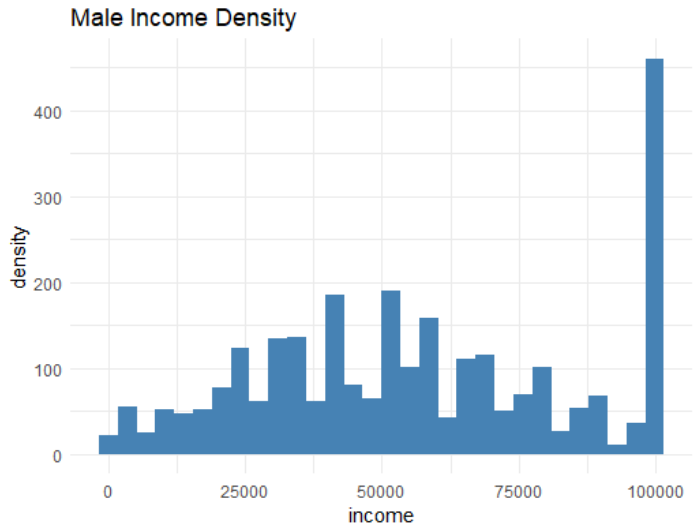
2. Education variables for parents are already continuous variables indicating years of schooling. I only change 95 to 0. Variable $YSCH - 3113$ indicating self education is translated to variable *edu*. Here provides the head of this variable.

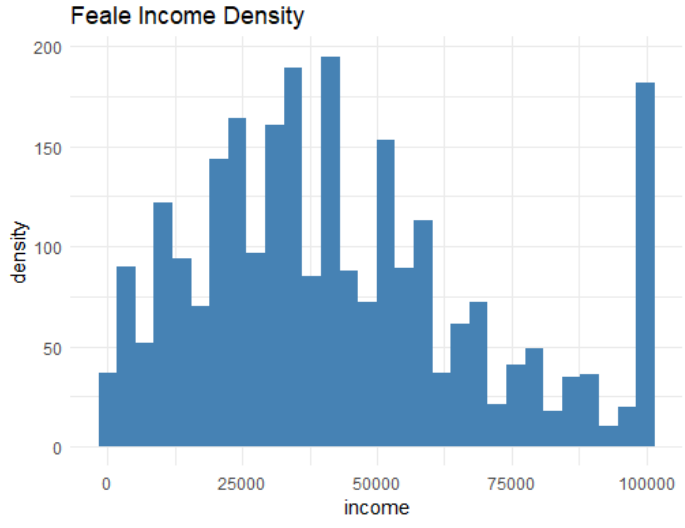
edu
NA
12
16
12
12
12

3. Here shows the average income by different age groups.The average income difference between different age groups is not large because their ages do not vary much.

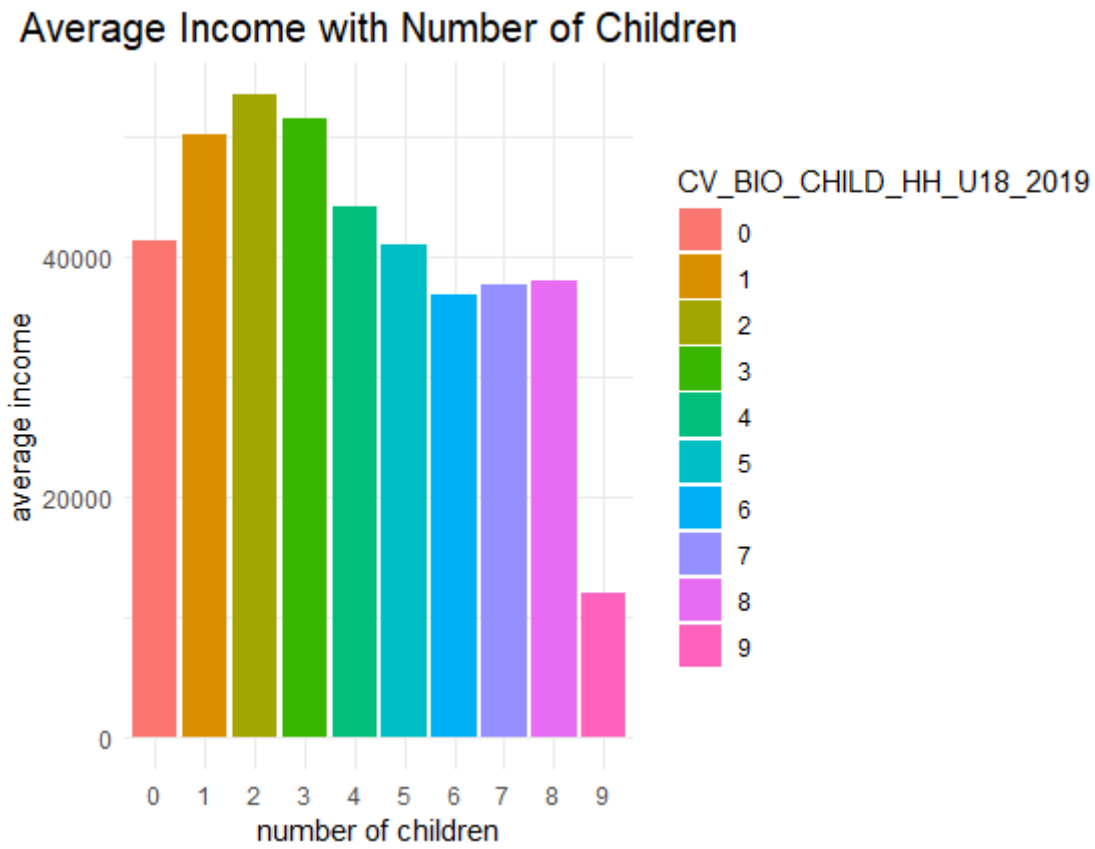


4. Here shows the average income by different age groups (1: male; 2: female), histogram of male's income, and histogram of female's income. Male's average income is higher than that of female. The histogram also shows that there's a large proportion of men whose incomes are top coded as 100000 while there are a large proportion of women whose incomes are below $50000.

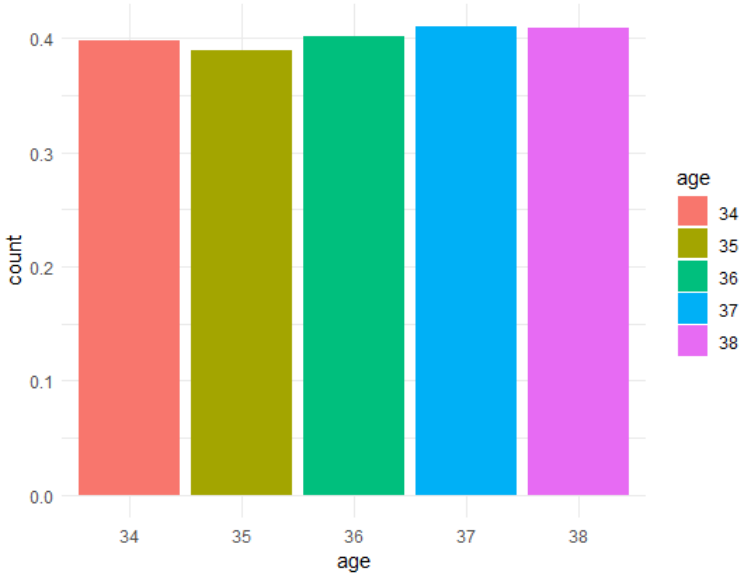




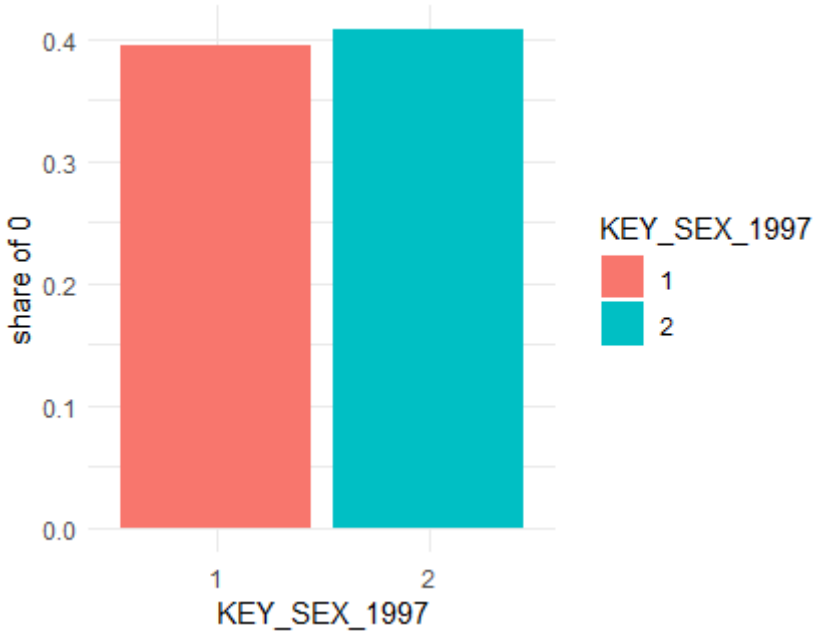
5. Here shows the average income by number of children an individual has. As having more children, the average income increases initially (0 to 2) and then generally decreases afterwards.(more than 2). A possible explanation is that poor families have less incentives to raise a child, which explains why the average income increases at first. For the decrease part, there's evidence showing that less educated people (with less income) have more children.



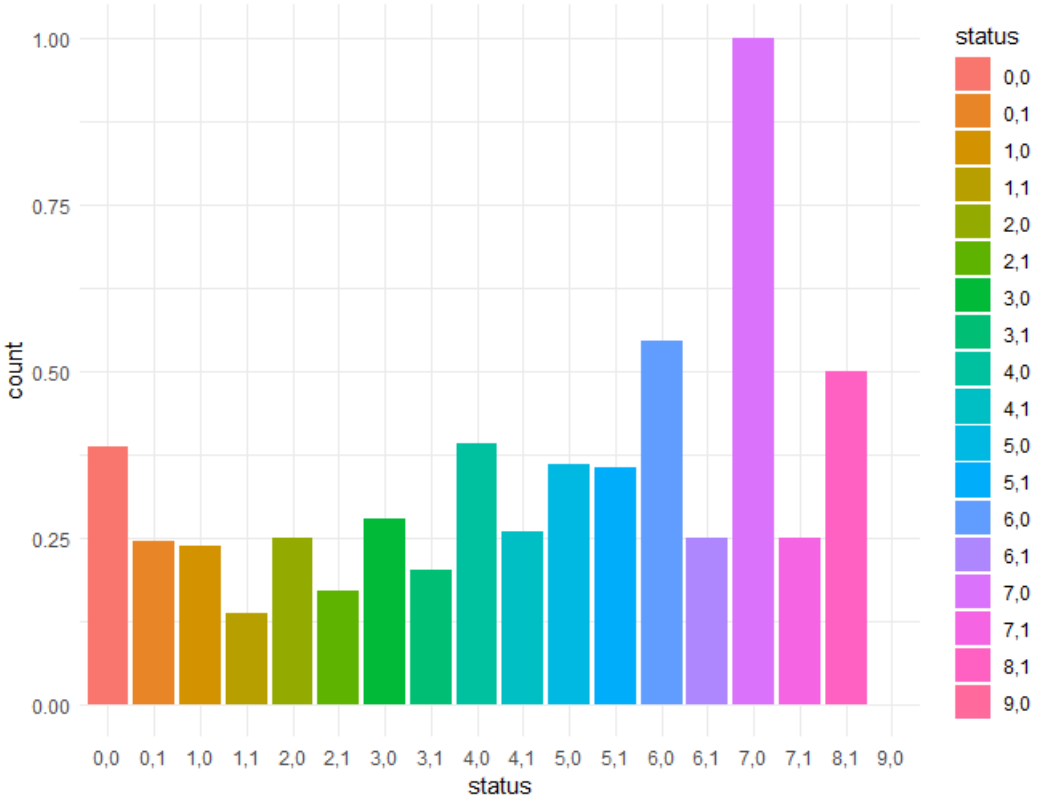
6. Here shows the share of 0 by age groups.The difference between different age groups is not large because their ages do not vary much.



7. Here shows the share of 0 by gender (1: male; 2: female). Women's share of 0 is slightly higher because housewives are more common than "househusband".



8. Here shows the share of 0 by the combination of marital status and number of children. I re-categorize the marital status variable to make it binary (0: single; 1: married).The status is a vector whose first element indicates marital status and the second element indicates number of children.



Exercise 2

1. I choose the natural log of income to be the dependent variable. I include age, gender, work experience, marital status, himself education and biological mother's education as independent variables. The result shows as follows.

```
Coefficients:
                           Estimate Std. Error t value             Pr(>|t|)
(Intercept)                9.942152   0.285854  34.781 < 0.0000000000000002 ***
age                       -0.001985   0.007827  -0.254              0.79977
KEY_SEX_1997              -0.371809   0.021879 -16.994 < 0.0000000000000002 ***
work_exp                   0.033423   0.002023  16.525 < 0.0000000000000002 ***
CV_MARSTAT_COLLAPSED_2019  0.129200   0.022230   5.812        0.00000000658 ***
edu                        0.063204   0.002901  21.788 < 0.0000000000000002 ***
CV_HGC_RES_MOM_1997        0.009101   0.003312   2.748              0.00601 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

   - All else equal, age does not have a significant effect on one's income.
   - All else equal, the women's average income is 37% less than that of men.
   - All else equal, one more year of work experience increases one's average income by approximately 3.3%.
   - All else equal, married people's average income is 12.9% higher than that of single people.
   - All else equal,one more year of schooling increases one's average income by approximately 6.3%.
   - All else equal,one more year of biological mother's schooling increases one's average income by approximately 0.9%.

   This regression omits people with 0 income and it is not random for one's income is 0. So there is sample selection bias. Heckman selection model takes the probability of one's income greater than 0 into consideration and then introduces the IV-like inverse mills ratio variable to correct the sample selection bias.

2. I use the same specification with the only difference is that there is one more variable *imr* (inverse mills ratio) in Heckman selection model. Here shows the second step regression result.

```
Coefficients:
                            Estimate Std. Error t value             Pr(>|t|)
(Intercept)                10.903435   0.289731  37.633 < 0.0000000000000002 ***
age                         0.009071   0.007730   1.173                0.241
KEY_SEX_1997               -0.206590   0.024777  -8.338 < 0.0000000000000002 ***
work_exp                   -0.022769   0.004644  -4.903      0.0000009766647 ***
edu                         0.026511   0.003953   6.706      0.0000000000223 ***
CV_MARSTAT_COLLAPSED_2019 -0.004101   0.023993  -0.171                0.864
CV_HGC_RES_MOM_1997         0.004057   0.003274   1.239                0.215
imr                        -1.839193   0.137407 -13.385 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

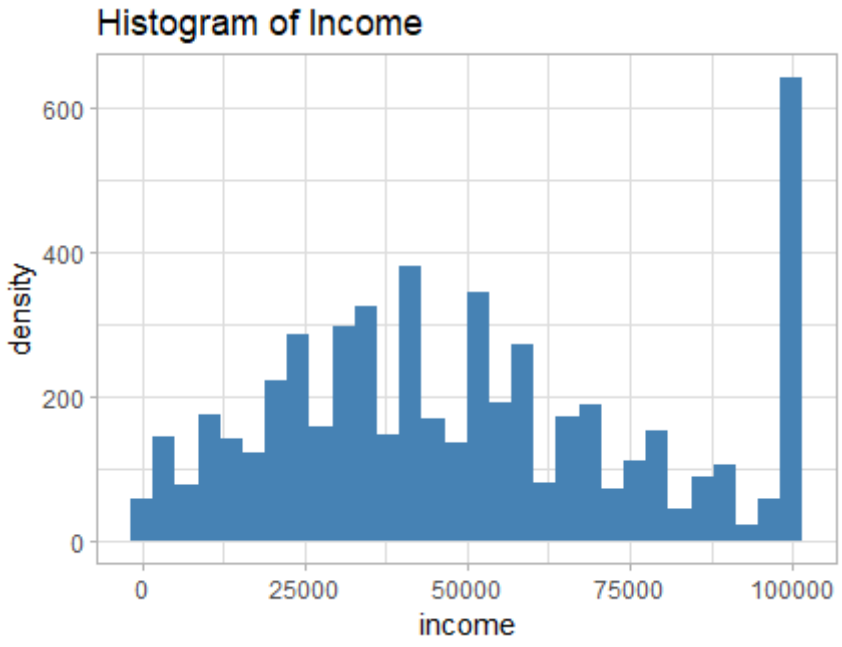
- All else equal, age does not have a significant effect on one's income.
- All else equal, the women's average income is 20.7% less than that of men.
- All else equal, one more year of work experience decreases one's average income by approximately 2.27%.
- All else equal,one more year of schooling increases one's average income by approximately 2.65%.
- All else equal, marital status does not have a significant effect on one's income.
- All else equal,biological mother's education does not have a significant effect on one's income.

The result is different because the sample selection bias is fixed.

Exercise 3

I use the data with income higher than 0 for this exercise.

1. Income is right censored at 100000.



2. I use the Tobit model and include variables: age, gender, experience, himself education, and marital status.Here shows the result of the Tobit model.

Variable	Coefficient
Intercept	23474.39288
Age	-17937.20554
Gender	303.92933
work_exp	961.81448
edu	8785.31189
mstatus	6762.4119
sigma	10.19464

- All else equal, a year increase in age decreases one's average income by approximately $17937.
- All else equal, the women's average income is not statistically different than that of men.
- All else equal, one more year of work experience increases one's average income by approximately $961.
- All else equal,one more year of schooling increases one's average income by approximately $8785.
- All else equal, married people's average income is approximately $6762 higher than that of single people.
- All else equal,biological mother's education does not have a significant effect on one's income.

3. Here provide the result when not correcting for the censored data. o_x_2 to o_x_6 represents age, gender, experience, himself education, and marital status respectively. Coefficients are different from those of Tobit model.

```
Coefficients:
             Estimate Std. Error t value             Pr(>|t|)
(Intercept)  27586.40    9044.50   3.050               0.0023 **
o_x2        -16065.41     700.23 -22.943 <0.0000000000000002 ***
o_x3           200.10     250.06   0.800               0.4236
o_x4           906.68      64.65  14.025 <0.0000000000000002 ***
o_x5          7768.45     240.13  32.350 <0.0000000000000002 ***
o_x6          5920.34     713.81   8.294 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exercise 4

1. The ability bias is that the most productive individuals also have an interest in studying for the longest period.

2. I show the between estimator first.

```
Coefficients:
                   Estimate Std. Error t-value             Pr(>|t|)
(Intercept)          1604.971    486.053   3.302            0.0009637 ***
edu                  1189.680     40.882  29.101 < 0.00000000000000022 ***
work_exp             2129.605     76.190  27.951 < 0.00000000000000022 ***
CV_MARSTAT_COLLAPSED 8825.639    567.109  15.562 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- All else equal, a year increase in education increases one's average income by approximately $1604.

- All else equal, one more year of work experience increases one's average income by approximately $2129.
- All else equal, married people's average income is approximately $8825 higher than that of single people.

3. Here shows the within estimator.

```
Coefficients:
                        Estimate           Std. Error t value            Pr(>|t|)
(Intercept)      0.00000000001188    69.71801462148782    0.00                   1
data$d_edu    1614.95129165004619    22.28813129062035   72.46 <0.0000000000000002 ***
data$d_exper  2821.13445972914451    26.24113684616926  107.51 <0.0000000000000002 ***
data$d_ms    15381.99887819730975   220.76033746128786   69.68 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- All else equal, a year increase in education increases one's average income by approximately $1614.
- All else equal, one more year of work experience increases one's average income by approximately $2821.
- All else equal, married people's average income is approximately $15381 higher than that of single people.

4. Here shows the first-difference estimator.

```
Coefficients:
                  Estimate Std. Error t value             Pr(>|t|)
(Intercept)       4065.65      68.34  59.488 < 0.0000000000000002 ***
data_43$fd_edu      90.44      22.02   4.108              0.00004 ***
data_43$fd_exper   954.77      29.68  32.169 < 0.0000000000000002 ***
data_43$fd_ms     2353.47     225.02  10.459 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- All else equal, a year increase in education increases one's average income by approximately $90.
- All else equal, one more year of work experience increases one's average income by approximately $954.
- All else equal, married people's average income is approximately $2353 higher than that of single people.

5. Different models yield different different parameter estimates because they process data in different ways. Between estimator discards time variation; within estimator discards individual variation; and first difference estimator preserves both variation.