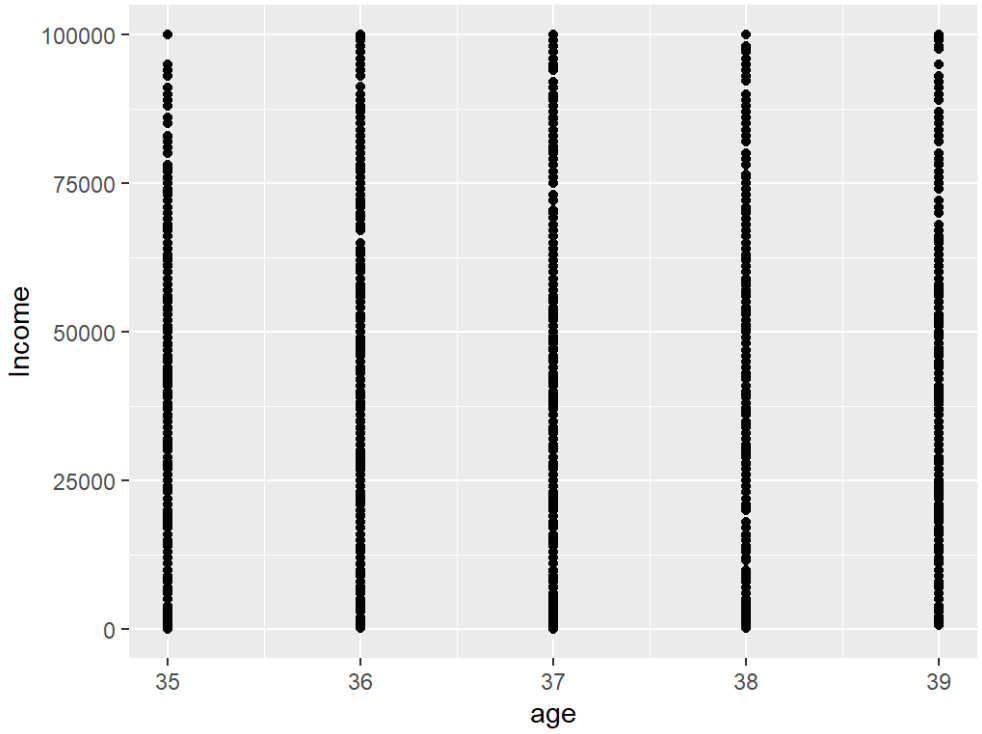
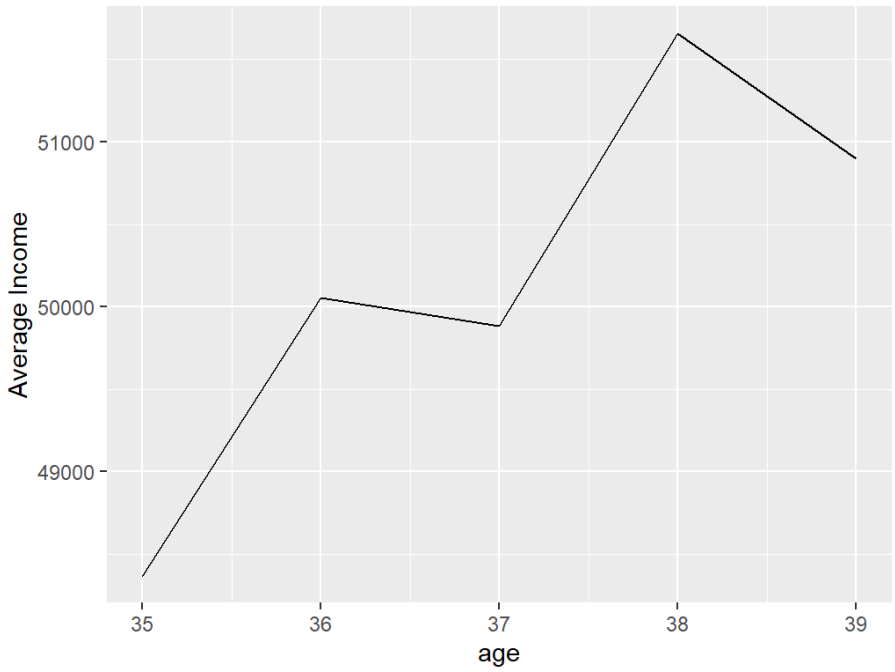
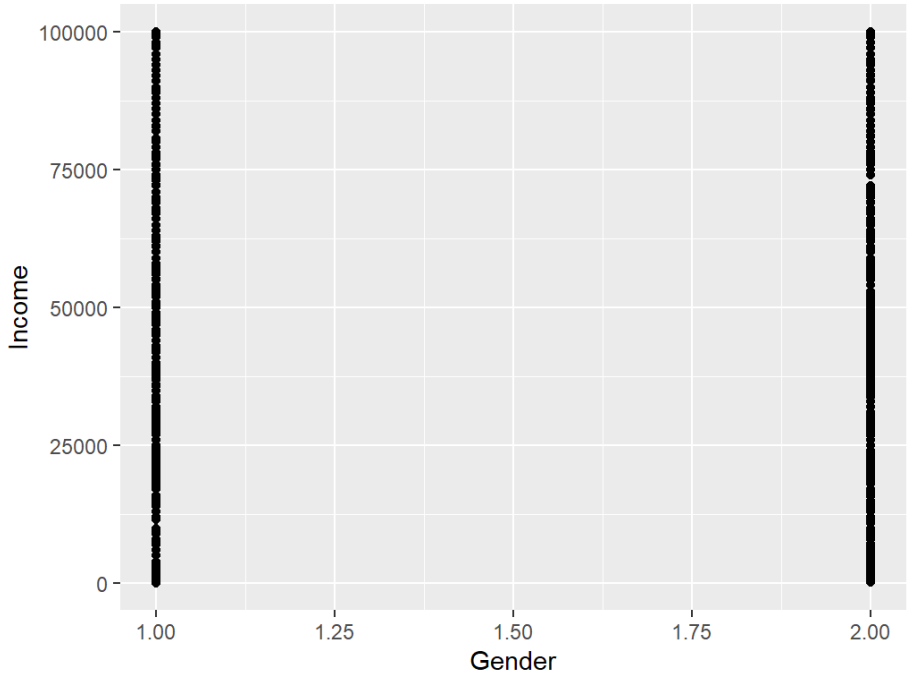
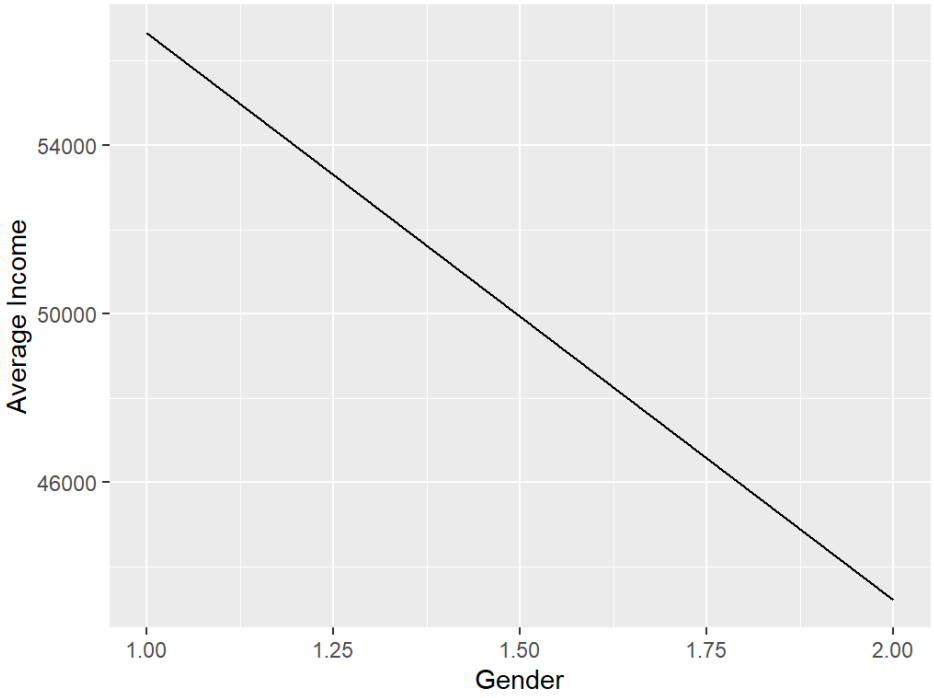
Problem 1

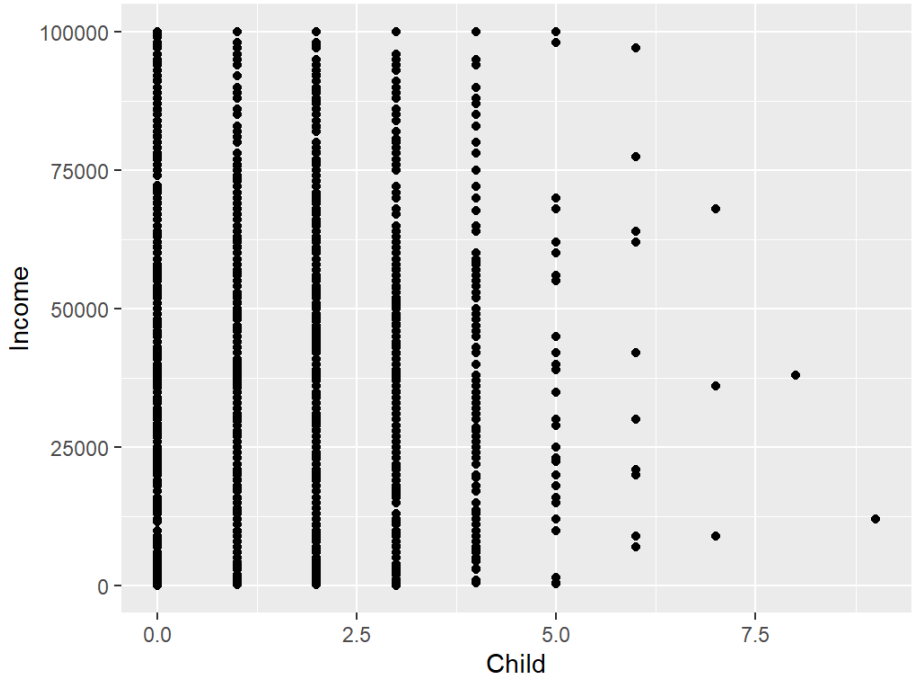
(c)

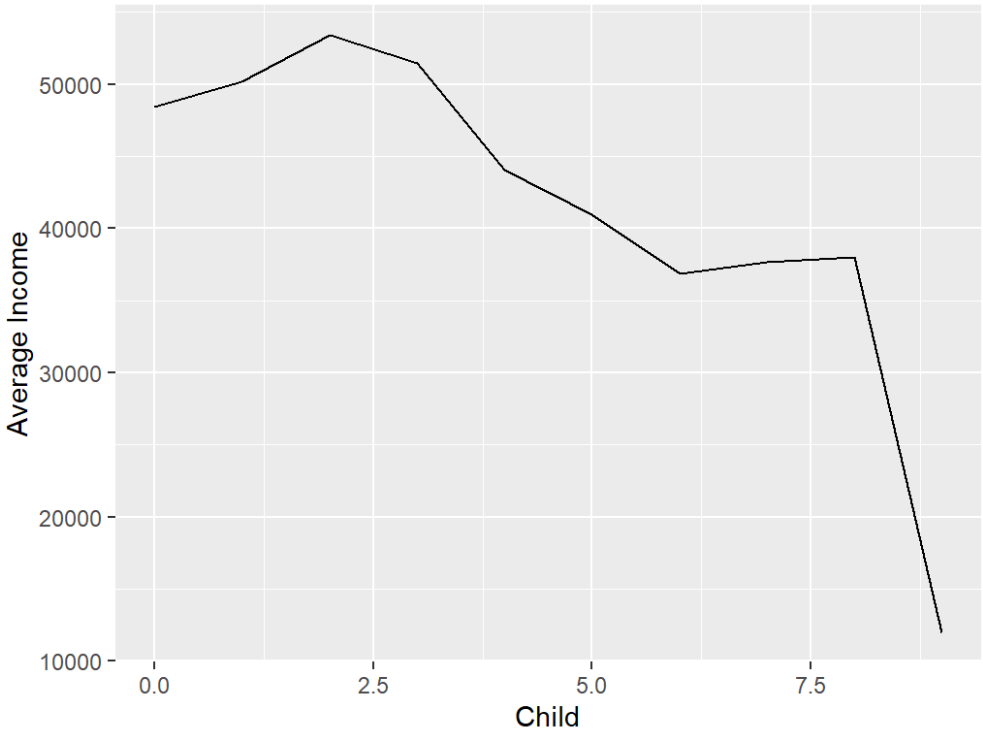

100000
75000
50000
25000
0
Income
35
36
37
38
39
age


51000
50000
49000
Average Income
35
36
37
38
39
age


100000
75000
50000
25000
0
Income
1.00
1.25
1.50
1.75
2.00
Gender


54000
50000
46000
Average Income
1.00
1.25
1.50
1.75
2.00
Gender


100000
75000
50000
25000
0
Income
0.0
2.5
5.0
7.5
Child


50000
40000
30000
20000
10000
Average Income
0.0
2.5
5.0
7.5
Child

	age	`YINC_1700_2019 == 0`	n
	<dbl>	<lgl>	<int>
1	35	FALSE	1066
2	35	TRUE	705
3	36	FALSE	1104
4	36	TRUE	703
5	37	FALSE	1101
6	37	TRUE	740
7	38	FALSE	1106
8	38	TRUE	768
9	39	FALSE	999
10	39	TRUE	692

	KEY_SEX_1997	`YINC_1700_2019 == 0`	n
	<dbl>	<lgl>	<int>
1	1	FALSE	2779
2	1	TRUE	1820
3	2	FALSE	2597
4	2	TRUE	1788

	CV_BIO_CHILD_HH_U18_2019	`YINC_1700_2019 == 0`	n
	<dbl>	<lgl>	<int>
1	0	FALSE	1988
2	0	TRUE	2696
3	1	FALSE	1138
4	1	TRUE	266
5	2	FALSE	1385
6	2	TRUE	343
7	3	FALSE	618
8	3	TRUE	183
9	4	FALSE	183
10	4	TRUE	82
11	5	FALSE	45
12	5	TRUE	25
13	6	FALSE	14
14	6	TRUE	10
15	7	FALSE	3
16	7	TRUE	2
17	8	FALSE	1
18	8	TRUE	1
19	9	FALSE	1

```
> 
```

Problem 2:

(a)

```
Call:
lm(formula = log(income) ~ Gender + biodad + biomom + resdad +
    resmom + hhsize + marstat + child + urbanrural + yearssch +
    age + work_exp + black + hispanic + nonbnonh, data = posinc)

Residuals:
    Min      1Q  Median      3Q     Max
-5.5486 -0.2566  0.1248  0.4510  1.8808

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.8686286  0.2980580  29.755  < 2e-16 ***
Gender      -0.4129094  0.0211414 -19.531  < 2e-16 ***
biodad       0.0006353  0.0022445   0.283  0.77714
biomom       0.0030392  0.0037020   0.821  0.41170
resdad       0.0028173  0.0018457   1.526  0.12696
resmom       0.0032969  0.0032120   1.026  0.30473
hhsize      -0.0524151  0.0098494  -5.322 1.07e-07 ***
marstat      0.0391630  0.0110955   3.530  0.00042 ***
child        0.0695484  0.0127932   5.436 5.68e-08 ***
urbanrural   0.1346791  0.0254952   5.283 1.32e-07 ***
yearssch     0.0979461  0.0038117  25.696  < 2e-16 ***
age          0.0048824  0.0073897   0.661  0.50883
work_exp     0.0337307  0.0019249  17.523  < 2e-16 ***
black       -0.1095982  0.1104869  -0.992  0.32126
hispanic     0.0460811  0.1106574   0.416  0.67711
nonbnonh     0.0069920  0.1095313   0.064  0.94910
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7494 on 5360 degrees of freedom
Multiple R-squared:  0.2437,    Adjusted R-squared:  0.2416
F-statistic: 115.1 on 15 and 5360 DF,  p-value: < 2.2e-16
```

(c)

```
Call:
glm(formula = inlf ~ Gender + biodad + biomom + resdad + resmom +
    hhsize + marstat + child + urbanrural + yearssch + age +
    work_exp + black + hispanic + nonbnonh, family = binomial(link = "probi
t"),
    data = dat_A4)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9930   0.1111   0.4499   0.7380   1.6437

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.6934906  0.5308406  -1.306 0.191416
Gender      -0.2849209  0.0386732  -7.367 1.74e-13 ***
biodad       0.0132585  0.0038303   3.462 0.000537 ***
biomom       0.0001498  0.0058949   0.025 0.979730
resdad      -0.0051748  0.0033518  -1.544 0.122616
resmom      -0.0012037  0.0052882  -0.228 0.819934
hhsize      -0.0050173  0.0162398  -0.309 0.757359
marstat      0.0798707  0.0193745   4.122 3.75e-05 ***
child        0.0205641  0.0213982   0.961 0.336542
urbanrural   0.0276423  0.0452841   0.610 0.541583
yearssch     0.0851288  0.0068540  12.420  < 2e-16 ***
age         -0.0074769  0.0131342  -0.569 0.569175
work_exp     0.1114958  0.0047744  23.353  < 2e-16 ***
black       -0.0747970  0.1965287  -0.381 0.703507
hispanic     0.1219491  0.1977375   0.617 0.537418
nonbnonh     0.1067581  0.1958305   0.545 0.585647
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7427.3  on 6946  degrees of freedom
Residual deviance: 6069.6  on 6931  degrees of freedom
AIC: 6101.6
```

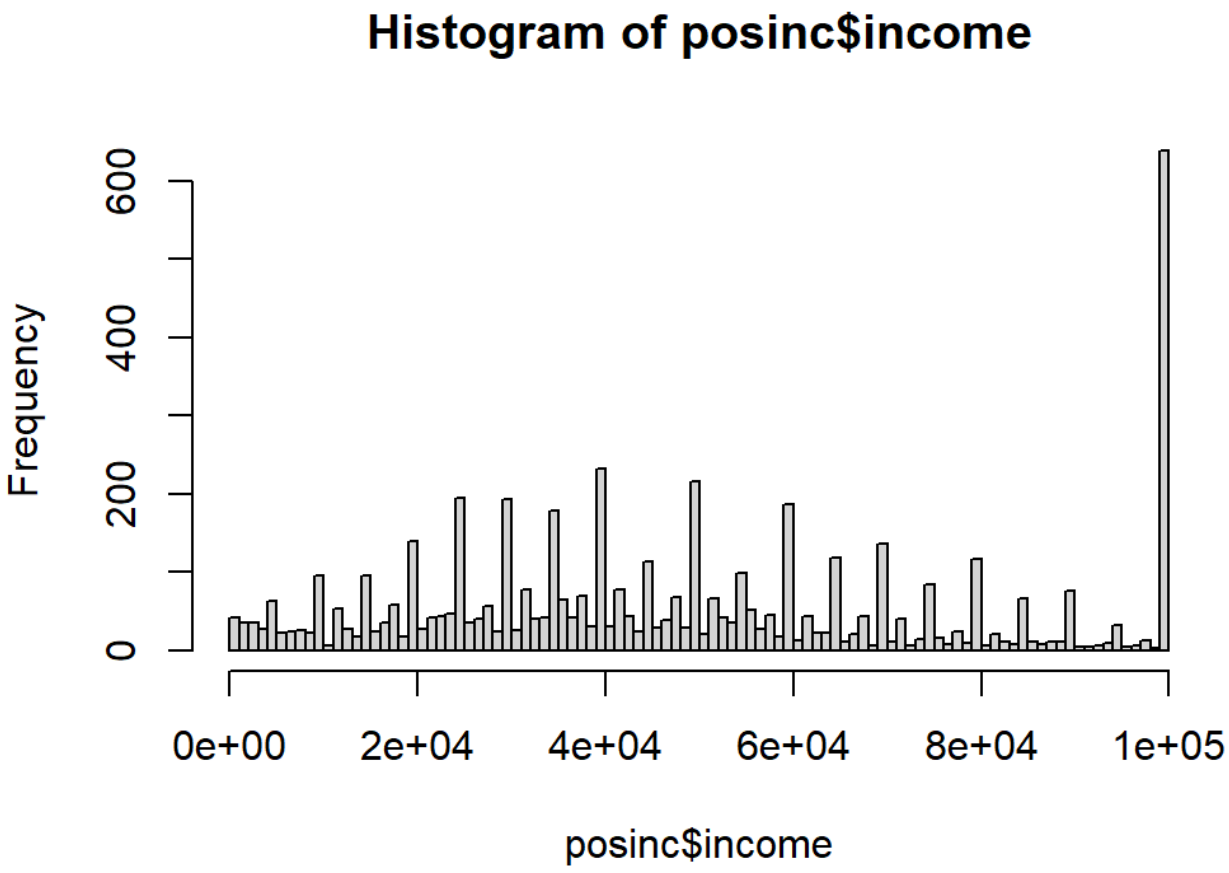
This is the estimation of the probability of a person participate in the labor market.

Second Stage:

```
Call:
lm(formula = log(income) ~ IMR + Gender + biodad + biomom + resdad +
    resmom + hhsize + marstat + child + urbanrural + yearssch +
    age + work_exp + black + hispanic + nonbnonh, data = posinc)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4448 -0.2530  0.1236  0.4332  2.4475

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.528215   0.321893  32.707  < 2e-16 ***
IMR         -1.324523   0.105036 -12.610  < 2e-16 ***
Gender      -0.268609   0.023772 -11.299  < 2e-16 ***
biodad      -0.003535   0.002237  -1.580 0.114087
biomom       0.001310   0.003651   0.359 0.719780
resdad       0.006714   0.001845   3.639 0.000276 ***
resmom       0.006381   0.003175   2.010 0.044517 *
hhsize      -0.086343   0.010073  -8.571  < 2e-16 ***
marstat     -0.016917   0.011805  -1.433 0.151922
child        0.084742   0.012666   6.690 2.45e-11 ***
urbanrural   0.049103   0.026028   1.887 0.059271 .
yearssch     0.028903   0.006640   4.353 1.37e-05 ***
age          0.009754   0.007293   1.337 0.181170
work_exp    -0.007907   0.003808  -2.076 0.037912 *
black       -0.094789   0.108900  -0.870 0.384108
hispanic     0.029241   0.109070   0.268 0.788635
nonbnonh    -0.037388   0.108009  -0.346 0.729238
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7386 on 5359 degrees of freedom
Multiple R-squared:  0.2655,    Adjusted R-squared:  0.2633
F-statistic: 121.1 on 16 and 5359 DF,  p-value: < 2.2e-16
```

Problem 3:

(a)



100,000 is the top coded value.

(b,c,d)

	Estimates
Gender	-0.484340492
biodad	0.001362190
biomom	0.003386790
resdad	0.004032734
resmom	0.007227052
hhsize	-0.031851519
marstat	0.032390773
child	0.068149624
urbanrural	0.220244260
yearssch	0.131366171
age	0.200516794
work_exp	0.032688270
black	0.954175976
hispanic	1.129718251
nonbnonh	1.092358518

This is the estimates after correcting the censoring problem.

Problem 4

Between estimator:

```
Call:
lm(formula = Y ~ educ + exp + mstat, data = within)

Residuals:
   Min     1Q Median     3Q    Max
-36037  -5421  -1093   3214  85446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -20335.65     605.34  -33.59   <2e-16 ***
educ          2493.81      59.38   42.00   <2e-16 ***
exp           2430.18      65.18   37.28   <2e-16 ***
mstat         6297.16     470.08   13.40   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9917 on 8980 degrees of freedom
Multiple R-squared:  0.417,     Adjusted R-squared:  0.4168
F-statistic:  2141 on 3 and 8980 DF,  p-value: < 2.2e-16
```

Within Estimator:

```
Call:
lm(formula = Ywithin ~ eduwithinols + expwithinols + mstatwithinols,
    data = within)

Residuals:
    Min      1Q  Median      3Q     Max
-105335   -7105       0    5741  303070

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.934e-10  4.223e+01    0.00        1
eduwithinols    2.798e+03  2.394e+01  116.88   <2e-16 ***
expwithinols    2.632e+03  1.826e+01  144.15   <2e-16 ***
mstatwithinols  1.093e+04  1.485e+02   73.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16980 on 161708 degrees of freedom
Multiple R-squared:  0.349,     Adjusted R-squared:  0.349
F-statistic: 2.89e+04 on 3 and 161708 DF,  p-value: < 2.2e-16
```

Difference Estimator:

```
Call:
lm(formula = differenceinc ~ differenceedu + differenceexp +
    differencemstat, data = Difference)

Residuals:
    Min      1Q  Median      3Q     Max
-236160   -3780    -998    2661  344130

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       997.68      42.92   23.24   <2e-16 ***
differenceedu    1896.02      28.85   65.72   <2e-16 ***
differenceexp    1787.98      21.50   83.16   <2e-16 ***
differencemstat  3527.52     115.12   30.64   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16550 on 152724 degrees of freedom
Multiple R-squared:  0.1253,    Adjusted R-squared:  0.1253
F-statistic:  7293 on 3 and 152724 DF,  p-value: < 2.2e-16
```