HW4

Mike Sun

4/20/2022

```
dat_A4 <- read.csv("~/Desktop/HW4/Data/dat_A4.csv")
```

*Exercise*1

Age & Work Years:

```
dat_A4 = dat_A4 %>%
  mutate(age = 2019-KEY_BDATE_Y_1997) %>%
  rowwise() %>%
  mutate(work_exp = sum(CV_WKSWK_JOB_DLI.01_2019,
                        CV_WKSWK_JOB_DLI.02_2019,
                        CV_WKSWK_JOB_DLI.03_2019,
                        CV_WKSWK_JOB_DLI.04_2019,
                        CV_WKSWK_JOB_DLI.05_2019,
                        CV_WKSWK_JOB_DLI.06_2019,
                        CV_WKSWK_JOB_DLI.07_2019,
                        CV_WKSWK_JOB_DLI.08_2019,
                        CV_WKSWK_JOB_DLI.09_2019,
                        CV_WKSWK_JOB_DLI.10_2019,
                        CV_WKSWK_JOB_DLI.11_2019,
                        na.rm=TRUE)/52)
```

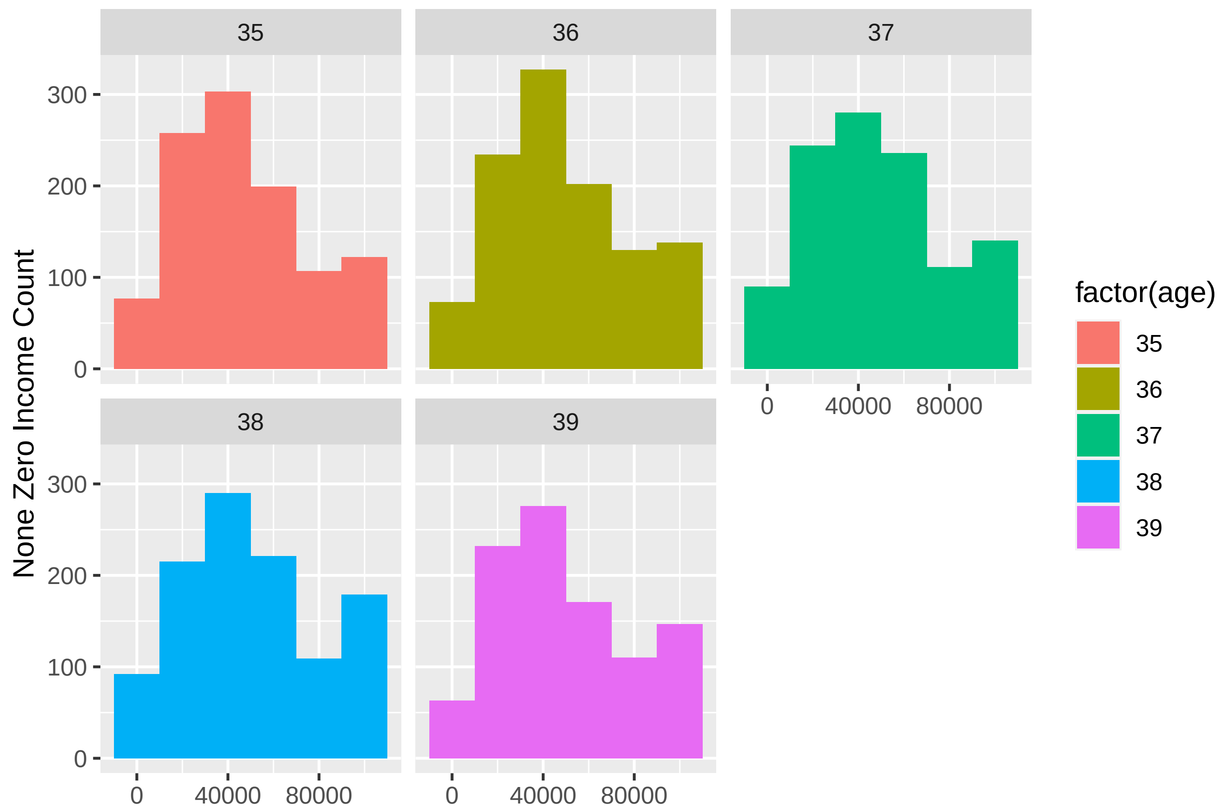
Education

```
dat_A4 = dat_A4 %>%
  rowwise() %>%
  mutate(CV_HGC_BIO_DAD_1997 = ifelse(CV_HGC_BIO_DAD_1997 >20,0,CV_HGC_BIO_DAD_1997)) %>%
  mutate(CV_HGC_BIO_MOM_1997 = ifelse(CV_HGC_BIO_MOM_1997 >20,0,CV_HGC_BIO_MOM_1997)) %>%
  mutate(CV_HGC_RES_DAD_1997 = ifelse(CV_HGC_RES_DAD_1997 >20,0,CV_HGC_RES_DAD_1997)) %>%
  mutate(CV_HGC_RES_MOM_1997 = ifelse(CV_HGC_RES_MOM_1997 >20,0,CV_HGC_RES_MOM_1997)) %>%
  mutate(edu = sum(CV_HGC_BIO_DAD_1997,
                   CV_HGC_BIO_MOM_1997,
                   CV_HGC_RES_DAD_1997,
                   CV_HGC_RES_MOM_1997,na.rm = TRUE))
```

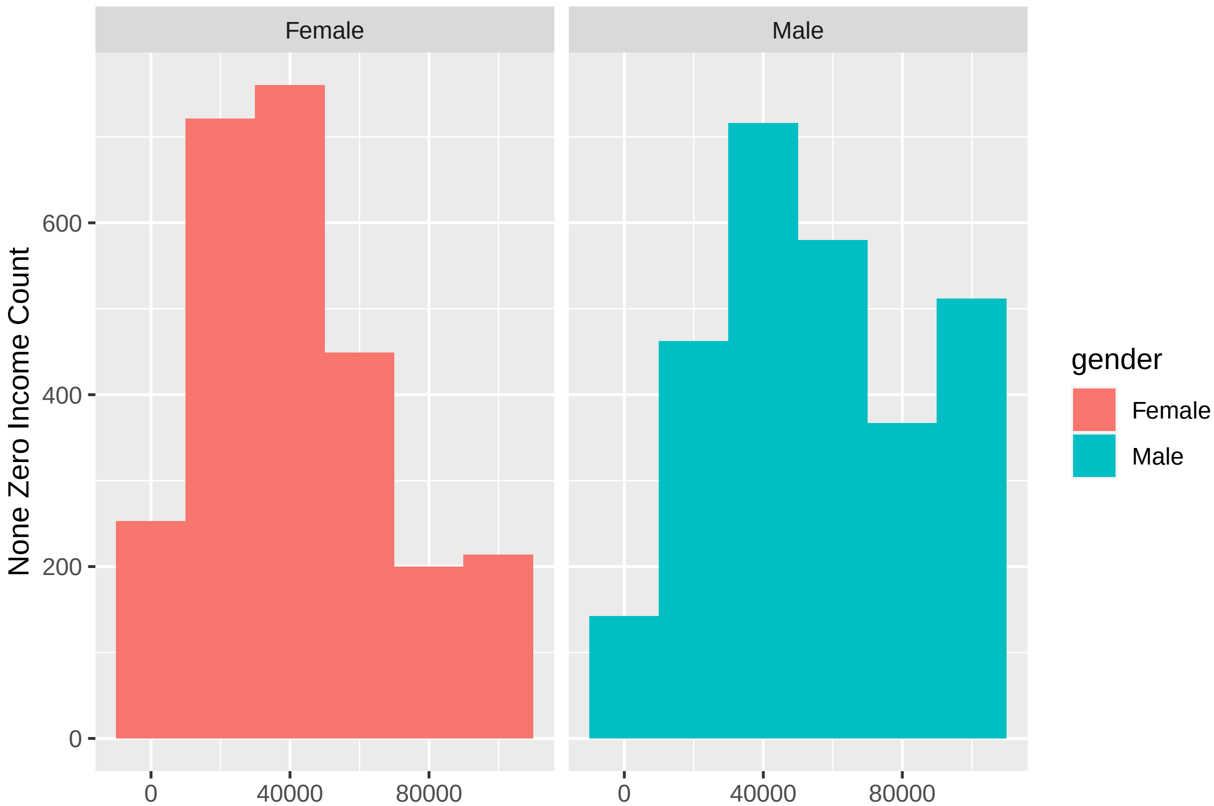
Positive income data by age & gender & number of children:

```
dat_A4 = dat_A4 %>%
  mutate(gender = ifelse(KEY_SEX_1997 == 1, "Male", ifelse(KEY_SEX_1997 == 2, "Female", "Unknown")))

dat_A4$gender = as.factor(dat_A4$gender)
dat_A4$num_child = factor(dat_A4$CV_BIO_CHILD_HH_U18_2019)

inc_age = dat_A4 %>% filter(YINC_1700_2019>0,na.rm =TRUE) %>%
```

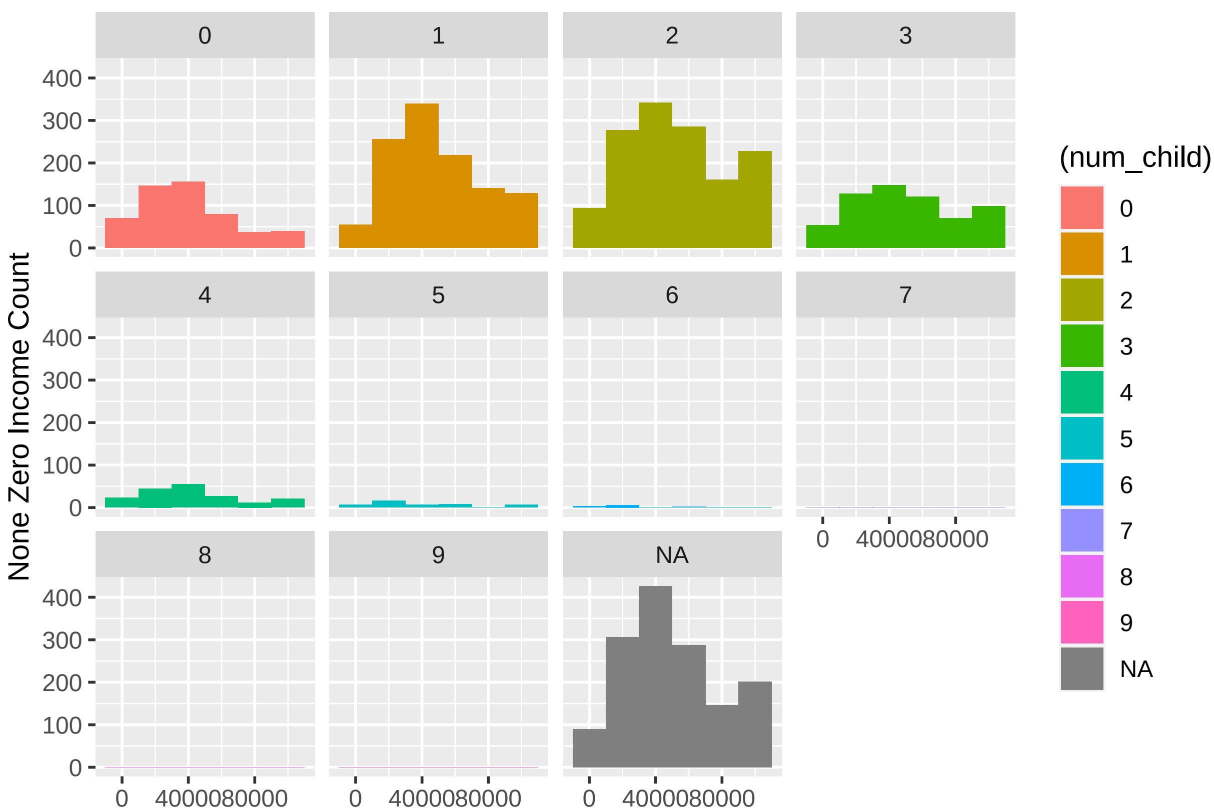
```r
  ggplot(aes(x=YINC_1700_2019, fill=factor(age),na.rm = TRUE)) +
  geom_histogram(position="identity", binwidth = 20000)+
  xlab("")+
  ylab("None Zero Income Count") +
  facet_wrap(~factor(age))

inc_gender = dat_A4 %>% filter(YINC_1700_2019>0) %>%
  ggplot(aes(x=YINC_1700_2019, fill=gender, na.rm = TRUE)) +
  geom_histogram(position="identity", binwidth = 20000)+
  xlab("")+
  ylab("None Zero Income Count") +
  facet_wrap(~factor(gender))

inc_chil = dat_A4 %>% filter(YINC_1700_2019>0) %>%
  ggplot(aes(x=YINC_1700_2019, fill=(num_child), na.rm = TRUE)) +
  geom_histogram(position="identity", binwidth = 20000)+
  xlab("")+
  ylab("None Zero Income Count") +
  facet_wrap(~factor(num_child))

inc_age
```




```r
inc_gender
```




```
inc_chil
```



0 income data share by group

```r
dat_gender_0_sum = dat_A4 %>%
  mutate(income_0 = ifelse(YINC_1700_2019 ==0,1,0)) %>%
  filter(income_0 == 1) %>%
  group_by(gender) %>%
  summarise(income_0_count = n()) %>%
  mutate(share_0 = income_0_count/sum(income_0_count))

dat_age_0_sum = dat_A4 %>%
  mutate(income_0 = ifelse(YINC_1700_2019 ==0,1,0)) %>%
  filter(income_0 == 1) %>%
  group_by(age) %>%
  summarise(income_0_count = n()) %>%
  mutate(share_0 = income_0_count/sum(income_0_count))

dat_chil_0_sum = dat_A4 %>%
  mutate(income_0 = ifelse(YINC_1700_2019 ==0,1,0)) %>%
  filter(income_0 == 1) %>%
  group_by(num_child = factor(CV_BIO_CHILD_HH_U18_2019)) %>%
  summarise(income_0_count = n()) %>%
  mutate(share_0 = income_0_count/sum(income_0_count))

dat_gender_0_sum
```

```
## # A tibble: 2 x 3
##   gender income_0_count share_0
##   <fct>           <int>   <dbl>
## 1 Female             15   0.417
## 2 Male               21   0.583
```

```r
dat_age_0_sum
```

```
## # A tibble: 5 x 3
##     age income_0_count share_0
##   <dbl>          <int>   <dbl>
## 1    35             10  0.278
## 2    36              7  0.194
## 3    37              6  0.167
## 4    38             10  0.278
## 5    39              3  0.0833
```

```r
dat_chil_0_sum
```

```
## # A tibble: 5 x 3
##   num_child income_0_count share_0
##   <fct>              <int>   <dbl>
## 1 0                      8   0.222
## 2 1                      9   0.25
## 3 2                      8   0.222
## 4 3                      5   0.139
## 5 <NA>                   6   0.167
```

Interpretation:

It turns out that age wise, those who are 38 at 2019 has more higher top income earner. However, there is less varaince across age group. Gender wise, Male has more higher income earner than female. This shows us that there may be gender pay gap. Lastly, those with one or two dependent child(ren) has more high income earner than those with none or more than three. This could lead to a hypothesis that healthy family structure (bearing one/two kid(s)) can lead to higher income. On the other hand, causality could be reversed: those with good income is more likely to build a family structure that has one or two children but not too many.

OLS:

```
OLS = lm(YINC_1700_2019 ~ age+work_exp+edu+gender+num_child, data=dat_A4)

summary(OLS)
```

```
##
## Call:
## lm(formula = YINC_1700_2019 ~ age + work_exp + edu + gender +
##     num_child, data = dat_A4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -73381 -18519  -2764  17409  81103
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -18672.91   10981.26  -1.700   0.0891 .
## age            676.70     292.07   2.317   0.0206 *
## work_exp      1150.13      77.46  14.849  < 2e-16 ***
## edu            362.77      25.30  14.339  < 2e-16 ***
## genderMale   18400.91     836.98  21.985  < 2e-16 ***
## num_child1   11272.47    1371.82   8.217 2.80e-16 ***
## num_child2   14640.52    1347.57  10.864  < 2e-16 ***
## num_child3   12952.74    1537.30   8.426  < 2e-16 ***
## num_child4    9033.01    2204.41   4.098 4.26e-05 ***
## num_child5    7601.84    3958.19   1.921   0.0549 .
## num_child6    2518.52    6882.23   0.366   0.7144
## num_child7   -5329.94   14723.05  -0.362   0.7174
## num_child8    8210.56   25432.69   0.323   0.7468
## num_child9  -25396.34   25421.72  -0.999   0.3179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25390 on 3933 degrees of freedom
##   (5037 observations deleted due to missingness)
## Multiple R-squared:  0.2297, Adjusted R-squared:  0.2272
## F-statistic: 90.22 on 13 and 3933 DF,  p-value: < 2.2e-16
```

The estimation results are consistent with the previous interpretion we have. Age has slight effect on income. Work experience and parents' education also improves the income. Being a male significantly increase income and having some number of children improves income whereas having too many would hurt it.

However, we face selection bias as there are too many NAs being excluded. After exclusion, there are only 36 people reporting 0 income, which is far below the natural unemployment rate given the remaining 3934 observations. Thus, we should worry that the sample we have is not representitive to the population.

The Heckman model can deal with the selection problem by finding the sampling probability for each observation such that we add additional controls for the likelihood of each observation's occurance to capture the selection bias problem. In this way, we first create a selection model which tells us if we observe the value of dependent variable for a person and the main model controlling for the fact that those unobeserved.

Heckman Selection Model:

```r
dat_A4 = dat_A4 %>% mutate(na_dummy = ifelse(is.na(YINC_1700_2019)==TRUE || YINC_1700_2019<0,1,0))
```

```r
selectionm = glm(na_dummy ~ age+work_exp+edu+gender+num_child,family=binomial(link="probit"),
                data=dat_A4)

dat_A4_nad = dat_A4 %>%
  filter(is.na(age) == FALSE) %>%
  filter(is.na(work_exp) == FALSE) %>%
  filter(is.na(edu) == FALSE) %>%
  filter(is.na(gender) == FALSE) %>%
  filter(is.na(num_child) == FALSE)

y_i = predict(selectionm, new_data = dat_A4$na_dummy)

dat_A4_nad$y_dum = 0

for (i in 1:length(dat_A4_nad$y_dum)){
  dat_A4_nad$y_dum[i]=y_i[i]
}

dat_A4_nad$IMR_na = dnorm(dat_A4_nad$y_dum)/pnorm(dat_A4_nad$y_dum)
```

```r
Heckman = lm(YINC_1700_2019 ~ age+work_exp+edu+gender+num_child+IMR_na, data=dat_A4_nad)

summary(Heckman)
```

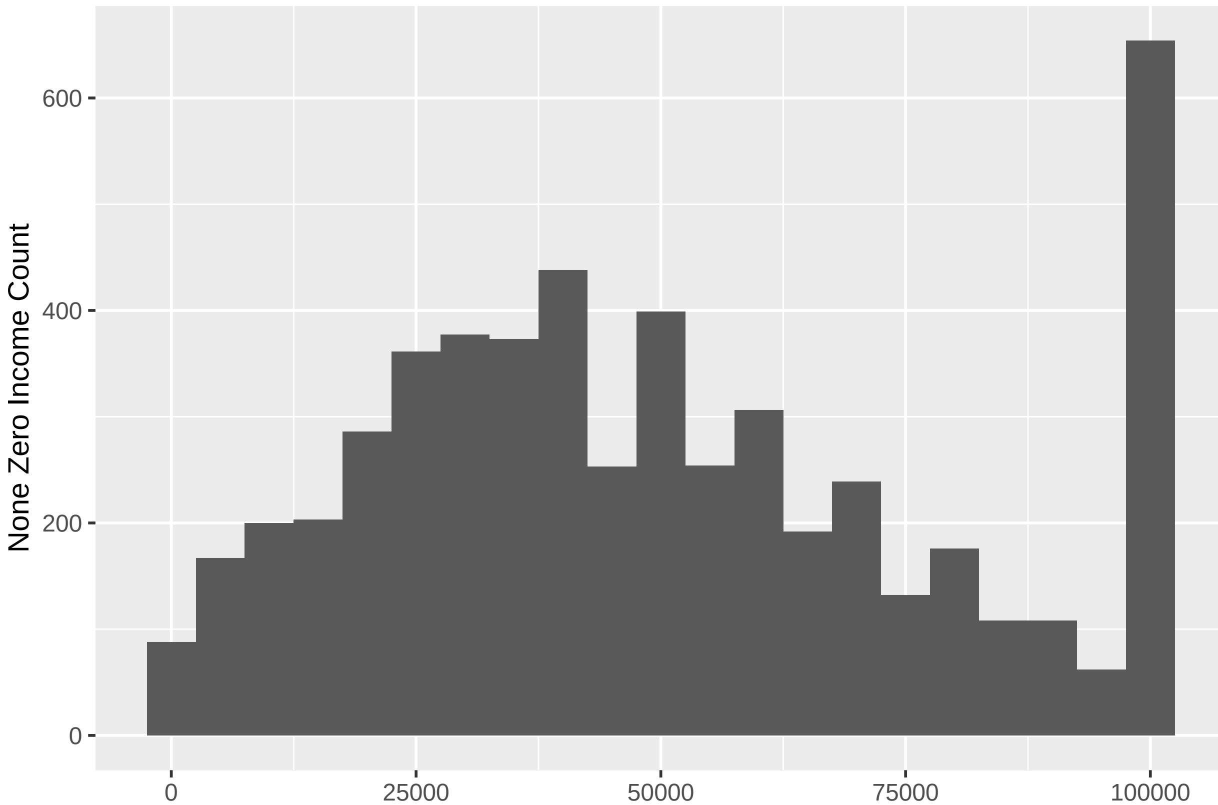
```
##
## Call:
## lm(formula = YINC_1700_2019 ~ age + work_exp + edu + gender +
##     num_child + IMR_na, data = dat_A4_nad)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -72024 -18325  -2795  17655  81902
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33248.89   11084.88  -2.999  0.00272 **
## age           1667.40     319.68   5.216 1.92e-07 ***
## work_exp     10776.37    1306.98   8.245 2.22e-16 ***
## edu            961.23      84.92  11.320  < 2e-16 ***
## genderMale   43399.36    3488.72  12.440  < 2e-16 ***
## num_child1   44001.42    4640.54   9.482  < 2e-16 ***
## num_child2   46930.40    4576.58  10.254  < 2e-16 ***
## num_child3   40707.12    4059.84  10.027  < 2e-16 ***
## num_child4   24390.78    3021.11   8.073 9.00e-16 ***
## num_child5   25121.12    4592.99   5.469 4.80e-08 ***
## num_child6    2914.42    6836.16   0.426  0.66990
## num_child7  -27985.81   14942.95  -1.873  0.06116 .
## num_child8   17810.02   25295.14   0.704  0.48142
## num_child9  331131.13   54522.28   6.073 1.37e-09 ***
## IMR_na      -93666.30   12695.25  -7.378 1.95e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 25220 on 3932 degrees of freedom
##   (1181 observations deleted due to missingness)
## Multiple R-squared:  0.2402, Adjusted R-squared:  0.2375
## F-statistic:  88.8 on 14 and 3932 DF,  p-value: < 2.2e-16
```

The Heckman model results show us that selection bias posed a very serious problem in previous OLS model. IMR variable here is very signifcant showing that nonrespondent would decrease income by a great degree. For other coefficient, age, work experience,edu, and Male each has more impact on income. The difference exist because we now assign probability weight on each observations and thus IMR here captures the potential selection bias.

Exercise3

```
inc = dat_A4 %>% filter(YINC_1700_2019>0,na.rm =TRUE) %>%
  ggplot(aes(x=YINC_1700_2019)) +
  geom_histogram(position="identity", binwidth = 5000)+
  xlab("")+
  ylab("None Zero Income Count")

inc
```



Based on the histogram, all income above $100,000 are censored.

The model I proposed to overcome this problem is still Heckman, but this time, I would use whether the data is censored or not as selection model upon the previous Heckman model.

```
dat_A4 = dat_A4 %>%
  mutate(censor_dummy = ifelse(YINC_1700_2019 != 100000 || is.na(YINC_1700_2019) == TRUE,0,1))
```

```
selectionm2 = glm(censor_dummy ~ age+work_exp+edu+gender+num_child, family=binomial(link="probit"),
                  data=dat_A4)

y_i2 = predict(selectionm2, new_data = dat_A4$censor_dummy)

dat_A4_nad$y_cdum = 0

for (i in 1:length(dat_A4_nad$y_cdum)){
  dat_A4_nad$y_cdum[i]=y_i2[i]
}
```

```r
dat_A4_nad$IMR_c = dnorm(dat_A4_nad$y_cdum)/pnorm(dat_A4_nad$y_cdum)
```

```r
Heckman2 = lm(YINC_1700_2019 ~ age+work_exp+edu+gender+num_child+IMR_c, data=dat_A4_nad)

summary(Heckman2)
```

```
##
## Call:
## lm(formula = YINC_1700_2019 ~ age + work_exp + edu + gender +
##     num_child + IMR_c, data = dat_A4_nad)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -81613 -18533  -2883  17203  78213
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -361541     108982  -3.317 0.000917 ***
## age             4283       1177   3.638 0.000278 ***
## work_exp        2933        569   5.154 2.67e-07 ***
## edu             1391        326   4.265 2.04e-05 ***
## genderMale     55236      11679   4.730 2.33e-06 ***
## num_child1     29499       5925   4.979 6.66e-07 ***
## num_child2     45594       9881   4.614 4.07e-06 ***
## num_child3     41429       9135   4.535 5.93e-06 ***
## num_child4     31391       7406   4.239 2.30e-05 ***
## num_child5     40245      11054   3.641 0.000275 ***
## num_child6   -221627      71216  -3.112 0.001871 **
## num_child7   -228056      71954  -3.169 0.001539 **
## num_child8   -159369      58770  -2.712 0.006722 **
## num_child9   -204585      62096  -3.295 0.000994 ***
## IMR_c          59165      18710   3.162 0.001578 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25360 on 3932 degrees of freedom
##   (1181 observations deleted due to missingness)
## Multiple R-squared:  0.2317, Adjusted R-squared:  0.2289
## F-statistic: 84.68 on 14 and 3932 DF,  p-value: < 2.2e-16
```

As we can see here, censoring problem is also very significantly impact our estimation. Removing censorship could potentially increase some of our estimates including age and education but show that work experience plays a less significant role. Each variables are more significantly positive after considering censorship problem and higher children number put a even higher burden on income, which is more reasonable.

*Exercise*4

```
dat_A4_panel <- read.csv("~/Desktop/HW4/Data/dat_A4_panel.csv")
```

Ability bias here captures the problem that those who are talented tend to get more education even though employers can also identity their talent and reward them with higher pay. Thus, there maybe overestimation of education on income.

```
dat_A4_panel <- dat_A4_panel %>% as_tibble()

work_year <- names(select(dat_A4_panel,contains("CV_WKSWK_JOB_DLI")))
```

```
dat_A4_panel <- dat_A4_panel %>%
  rowwise(X) %>%
  mutate(work_exp_1997 = sum(c_across(work_year[1]:work_year[7]),na.rm = TRUE)/52) %>%
  mutate(work_exp_1998 = sum(c_across(work_year[8]:work_year[16]),na.rm = TRUE)/52) %>%
  mutate(work_exp_1999 = sum(c_across(work_year[17]:work_year[25]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2000 = sum(c_across(work_year[26]:work_year[34]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2001 = sum(c_across(work_year[35]:work_year[42]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2002 = sum(c_across(work_year[43]:work_year[53]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2003 = sum(c_across(work_year[54]:work_year[63]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2004 = sum(c_across(work_year[64]:work_year[70]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2005 = sum(c_across(work_year[71]:work_year[79]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2006 = sum(c_across(work_year[80]:work_year[88]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2007 = sum(c_across(work_year[89]:work_year[96]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2008 = sum(c_across(work_year[97]:work_year[104]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2009 = sum(c_across(work_year[105]:work_year[113]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2010 = sum(c_across(work_year[114]:work_year[122]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2011 = sum(c_across(work_year[123]:work_year[135]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2013 = sum(c_across(work_year[136]:work_year[145]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2015 = sum(c_across(work_year[146]:work_year[157]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2017 = sum(c_across(work_year[158]:work_year[172]),na.rm = TRUE)/52) %>%
  mutate(work_exp_2019 = sum(c_across(work_year[173]:work_year[183]),na.rm = TRUE)/52)
```

```
dat_p_work = dat_A4_panel %>%
  select(X,
    work_exp_1997,
         work_exp_1998,
         work_exp_1999,
         work_exp_2000,
         work_exp_2001,
         work_exp_2002,
         work_exp_2003,
         work_exp_2004,
         work_exp_2005,
         work_exp_2006,
         work_exp_2007,
         work_exp_2008,
         work_exp_2009,
         work_exp_2010,
         work_exp_2011,
         work_exp_2013,
         work_exp_2015,
         work_exp_2017,
```

```
         work_exp_2019,
  )

dat_p_edu = dat_A4_panel %>%
  select(X, CV_HIGHEST_DEGREE_9899_1998,
         CV_HIGHEST_DEGREE_9900_1999,
         CV_HIGHEST_DEGREE_0001_2000,
         CV_HIGHEST_DEGREE_0102_2001,
         CV_HIGHEST_DEGREE_0203_2002,
         CV_HIGHEST_DEGREE_0304_2003,
         CV_HIGHEST_DEGREE_0405_2004,
         CV_HIGHEST_DEGREE_0506_2005,
         CV_HIGHEST_DEGREE_0607_2006,
         CV_HIGHEST_DEGREE_0708_2007,
         CV_HIGHEST_DEGREE_0809_2008,
         CV_HIGHEST_DEGREE_0910_2009,
         CV_HIGHEST_DEGREE_1011_2010,
         CV_HIGHEST_DEGREE_1112_2011,
         CV_HIGHEST_DEGREE_EVER_EDT_2013,
         CV_HIGHEST_DEGREE_EVER_EDT_2015,
         CV_HIGHEST_DEGREE_EVER_EDT_2017,
         CV_HIGHEST_DEGREE_EVER_EDT_2019)

dat_p_inc = dat_A4_panel %>%
  select(X,
         YINC.1700_1997,
         YINC.1700_1998,
         YINC.1700_1999,
         YINC.1700_2000,
         YINC.1700_2001,
         YINC.1700_2002,
         YINC.1700_2003,
         YINC.1700_2004,
         YINC.1700_2005,
         YINC.1700_2006,
         YINC.1700_2007,
         YINC.1700_2008,
         YINC.1700_2009,
         YINC.1700_2010,
         YINC.1700_2011,
         YINC.1700_2013,
         YINC.1700_2015,
         YINC.1700_2017,
         YINC.1700_2019)

dat_p_mars = dat_A4_panel %>%
  select(X,
         CV_MARSTAT_COLLAPSED_1997,
         CV_MARSTAT_COLLAPSED_1998,
         CV_MARSTAT_COLLAPSED_1999,
         CV_MARSTAT_COLLAPSED_2000,
         CV_MARSTAT_COLLAPSED_2001,
         CV_MARSTAT_COLLAPSED_2002,
```

```
         CV_MARSTAT_COLLAPSED_2003,
         CV_MARSTAT_COLLAPSED_2004,
         CV_MARSTAT_COLLAPSED_2005,
         CV_MARSTAT_COLLAPSED_2006,
         CV_MARSTAT_COLLAPSED_2007,
         CV_MARSTAT_COLLAPSED_2008,
         CV_MARSTAT_COLLAPSED_2009,
         CV_MARSTAT_COLLAPSED_2010,
         CV_MARSTAT_COLLAPSED_2011,
         CV_MARSTAT_COLLAPSED_2013,
         CV_MARSTAT_COLLAPSED_2015,
         CV_MARSTAT_COLLAPSED_2017,
         CV_MARSTAT_COLLAPSED_2019
         )
```

```
x = c("X","1997","1998","1999",
      "2000","2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","2011",
      "2013","2015","2017","2019")

colnames(dat_p_work) = x
colnames(dat_p_edu) = x
colnames(dat_p_inc) = x
colnames(dat_p_mars) = x

dat_work_pivot = dat_p_work %>%
  pivot_longer(!X,names_to = "year",values_to = "work_exp")

dat_edu_pivot = dat_p_edu %>%
   pivot_longer(!X,names_to = "year",values_to = "edu")

dat_inc_pivot = dat_p_inc %>%
   pivot_longer(!X,names_to = "year",values_to = "inc")

dat_edu_mars = dat_p_mars %>%
   pivot_longer(!X,names_to = "year",values_to = "marstats")
```

```
dat_pcleaned = left_join(dat_work_pivot,dat_edu_pivot,by=c("X","year"))
dat_pcleaned = left_join(dat_pcleaned,dat_inc_pivot,by=c("X","year"))
dat_pcleaned = left_join(dat_pcleaned,dat_edu_mars,by=c("X","year"))

dat_pcleaned_dis = unique(dat_pcleaned)

pd <- pdata.frame(dat_pcleaned_dis, index = c("X", "year")) %>% na.omit
```

```
panel_reg_within = plm(inc ~ work_exp + edu + marstats, data = pd,
    model = "within")

summary(panel_reg_within)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = inc ~ work_exp + edu + marstats, data = pd, model = "within")
##
## Unbalanced Panel: n = 8373, T = 1-18, N = 73069
```

```
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -124694.282   -7715.548     -93.492    6871.574  188036.150
##
## Coefficients:
##          Estimate Std. Error t-value  Pr(>|t|)
## work_exp 2511.333     27.970  89.787 < 2.2e-16 ***
## edu      9252.966     98.785  93.668 < 2.2e-16 ***
## marstats 7549.805    142.937  52.819 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3.1387e+13
## Residual Sum of Squares: 2.025e+13
## R-Squared:      0.35482
## Adj. R-Squared: 0.27129
## F-statistic: 11859.2 on 3 and 64693 DF, p-value: < 2.22e-16
```

```
panel_reg_between = plm(inc ~ work_exp + edu + marstats, data = pd,model = "between")

summary(panel_reg_between)
```

```
## Oneway (individual) effect Between Model
##
## Call:
## plm(formula = inc ~ work_exp + edu + marstats, data = pd, model = "between")
##
## Unbalanced Panel: n = 8373, T = 1-18, N = 73069
## Observations used in estimation: 8373
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -62877.5  -8128.4  -2165.1   5435.5 185213.5
##
## Coefficients:
##             Estimate Std. Error t-value  Pr(>|t|)
## (Intercept) 3565.428    350.427  10.175 < 2.2e-16 ***
## work_exp    2226.935     70.951  31.387 < 2.2e-16 ***
## edu         4565.819    131.150  34.814 < 2.2e-16 ***
## marstats    3522.882    280.656  12.552 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.0269e+12
## Residual Sum of Squares: 1.4896e+12
## R-Squared:      0.26511
## Adj. R-Squared: 0.26484
## F-statistic: 1006.35 on 3 and 8369 DF, p-value: < 2.22e-16
```

```
panel_reg_diff = plm(inc ~ work_exp + edu + marstats, data = pd,model = "fd")

summary(panel_reg_diff)
```

```
## Oneway (individual) effect First-Difference Model
```

```
##
## Call:
## plm(formula = inc ~ work_exp + edu + marstats, data = pd, model = "fd")
##
## Unbalanced Panel: n = 8373, T = 1-18, N = 73069
## Observations used in estimation: 64696
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -211048.1   -5364.1   -1950.0    4238.2  230439.6
##
## Coefficients:
##             Estimate Std. Error t-value  Pr(>|t|)
## (Intercept) 3631.572     64.072 56.6795 < 2.2e-16 ***
## work_exp     834.201     30.259 27.5686 < 2.2e-16 ***
## edu          761.097    118.121  6.4433 1.177e-10 ***
## marstats    1815.267    154.777 11.7282 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.4656e+13
## Residual Sum of Squares: 1.444e+13
## R-Squared:      0.014751
## Adj. R-Squared: 0.014705
## F-statistic: 322.848 on 3 and 64692 DF, p-value: < 2.22e-16
```

The results yield different estimates because within model is the fixed effect model, prespecified in individual and year. This model tells us the changes of on indiviudal level if estimators differ. Form the model, we can see that on average, if one has one additional year of work experience, his income is increased by 2511. However, for between model, we try to answer the question on the expected difference between two indiviudals if they differ on the explored independent variable. Thus, the interpretation is that if one has one additional work of experience, comparing to others, he can get 3565 increase in income. Lastly, fd model is estimating the first differentiation changes between dependent and independent variables one period before. Thus, its interpretation is that comparing to this and last year, increase work experience will increase income by 834 on average.