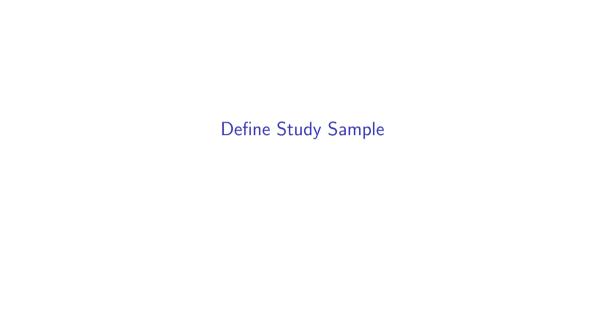
Deduction data

Last updated on 2023/11/17



Raw Data and Sample Year

```
raw <- read_csv(here("data/shaped2.csv"))</pre>
```

Observations of raw data: N = 167912

We define study sample, following conditions.

```
use1 <- subset(raw, 2010 <= year & year < 2018)
```

Condition 1: Observations in 2010–2017 (N = 110223)

Condition: Household Members

```
use2 <- subset(use1, !is.infinite(hh_max_inc))</pre>
```

Condition 2: No information on labor income of all household members (N=109436)

```
use3 <- subset(use2, family_position == 1 & work %in% c(1, 3))
```

Condition 3: Household headers who are full-time wage earner or self-employed (N=25659)

Condition: Age

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 19.0 40.0 48.0 48.7 56.0 93.0

use4 <- subset(use3, 24 <= age)</pre>
```

Condition 4: Those who are 24 years old or more (N = 25575)

Sample Characteristics: Income (1)

```
summary(use4$linc)
##
      Min. 1st Qu. Median
                              Mean 3rd Qu.
                                               Max.
                                                       NA's
##
         0
                 0
                      2114
                              2782
                                       4500
                                              66000
                                                          10
summary(use4$tinc)
##
      Min. 1st Qu. Median
                              Mean 3rd Qu.
                                               Max.
                                                       NA's
##
              2030
                      3571
                                       5500
         0
                              4207
                                              75300
                                                         99
```

Sample Characteristics: Income (2)

```
with(use4, table(linc == hh max inc, useNA = "always"))
##
## FALSE TRUE
                <NA>
   4088 21477
##
                  10
with(use4, table(linc == tinc, useNA = "always"))
##
## FALSE TRUE
                <NA>
## 13387 12089
                  99
```

Condition: Income

```
use5 <- subset(use4, linc == hh_max_inc)</pre>
```

Condition 5: Those who are the highest earners (N = 21477)

Condition 6: Labor income is less than 300 million KRW (N=21464)

Revisit: Sample Characteristics

```
summary(use6$linc)
##
     Min. 1st Qu. Median Mean 3rd Qu.
                                         {\tt Max.}
                           3157 4939 29872
##
        0
               0
                    2722
summary(use6$tinc)
##
     Min. 1st Qu. Median Mean 3rd Qu. Max.
                                                  NA's
##
        0
             2252
                    3600
                           4397
                                   5880
                                          75300
                                                    75
with(use6, table(linc == tinc, useNA = "always"))
##
## FALSE TRUE
               <NA>
  10136 11253
                75
```