# HelloWorld.rmd

2023-03-17

## the document is by R11323019 .

## 1.1

```
print("HelloWorld")
```

```
## [1] "HelloWorld"
```

## 1.3

```
add_two_number <- function(intA,intB ) {
    return(intA + intB)
}
add_two_number(2,3)
```

```
## [1] 5
```

## 3.4

```
# import data
library(haven)
```

```
## Warning:   'haven'   R   4.2.2
```

```
library('dplyr')
```

```
## Warning:   'dplyr'   R   4.2.2
```

```
##
##    'dplyr'

##       'package:stats':
##
##     filter, lag
```

```
##       'package:base':
##
##     intersect, setdiff, setequal, union
```

```
PSFD2016 <- read_sav("C:/Users/user/Downloads/C00320_3/RCI2016_v201811_spss.sav")

head(PSFD2016)
```

```
## # A tibble: 6 x 1,003
##     x01 x01a       x01b    x01c     x01d  x01e   x02 x03z01 x03z02 x03z03 x03z04
##   <dbl> <dbl+lbl>  <dbl+l> <dbl+l> <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 10021 3 [3  ~ 6 [6 ~ 1 [1 ~  1002     1  2016      2      3     13      6
## 2 10023 3 [3  ~ 6 [6 ~ 1 [1 ~  1002     3  2016      2      3     18     45
## 3 10121 3 [3  ~ 6 [6 ~ 1 [1 ~  1012     1  2016      1     12     21     53
## 4 10251 3 [3  ~ 6 [6 ~ 1 [1 ~  1025     1  2016      2     15     11     33
## 5 10252 3 [3  ~ 6 [6 ~ 1 [1 ~  1025     2  2016      2     15     12     28
## 6 10382 3 [3  ~ 6 [6 ~ 1 [1 ~  1038     2  2016      3      4     16     26
## # ... with 992 more variables: x05 <dbl+lbl>, x07 <dbl+lbl>, a01 <dbl+lbl>,
## #   a02a01 <dbl+lbl>, a02a02 <dbl+lbl>, a02b <dbl+lbl>, a03a01 <dbl+lbl>,
## #   a03a02 <dbl+lbl>, a03a03 <dbl+lbl>, a03b <dbl+lbl>, a04a <dbl+lbl>,
## #   a04b <dbl+lbl>, a04c <dbl+lbl>, a04d <dbl+lbl>, a05a <dbl+lbl>,
## #   a05b <dbl+lbl>, a05c <dbl+lbl>, a05d <dbl+lbl>, a06 <dbl+lbl>,
## #   a07 <dbl+lbl>, a08 <dbl+lbl>, b01 <dbl+lbl>, b02z01 <dbl+lbl>,
## #   b02z02 <dbl+lbl>, b03a01 <dbl+lbl>, b03a02 <dbl+lbl>, b03a03 <dbl+lbl>, ...
```

```
#check if   (x02) == 2016
PSFD2016[PSFD2016$x02 != 2016,]
```

```
## # A tibble: 0 x 1,003
## # ... with 1,003 variables: x01 <dbl>, x01a <dbl+lbl>, x01b <dbl+lbl>,
## #   x01c <dbl+lbl>, x01d <dbl>, x01e <dbl>, x02 <dbl>, x03z01 <dbl>,
## #   x03z02 <dbl>, x03z03 <dbl>, x03z04 <dbl>, x05 <dbl+lbl>, x07 <dbl+lbl>,
## #   a01 <dbl+lbl>, a02a01 <dbl+lbl>, a02a02 <dbl+lbl>, a02b <dbl+lbl>,
## #   a03a01 <dbl+lbl>, a03a02 <dbl+lbl>, a03a03 <dbl+lbl>, a03b <dbl+lbl>,
## #   a04a <dbl+lbl>, a04b <dbl+lbl>, a04c <dbl+lbl>, a04d <dbl+lbl>,
## #   a05a <dbl+lbl>, a05b <dbl+lbl>, a05c <dbl+lbl>, a05d <dbl+lbl>, ...
```

```
#calculate age
##a02a01:
##  =2016 =  105 )

PSFD2016$age <- 105 - PSFD2016$a02a01 + 1

#c02z01:     ? 1:  2 3:   4:
##only consider work at market(1).
# %>% mutate(work = if_else(c02z01 == 1, 1, 0))
PSFD2016$work <- as.integer(PSFD2016$c02z01 == 1)
```

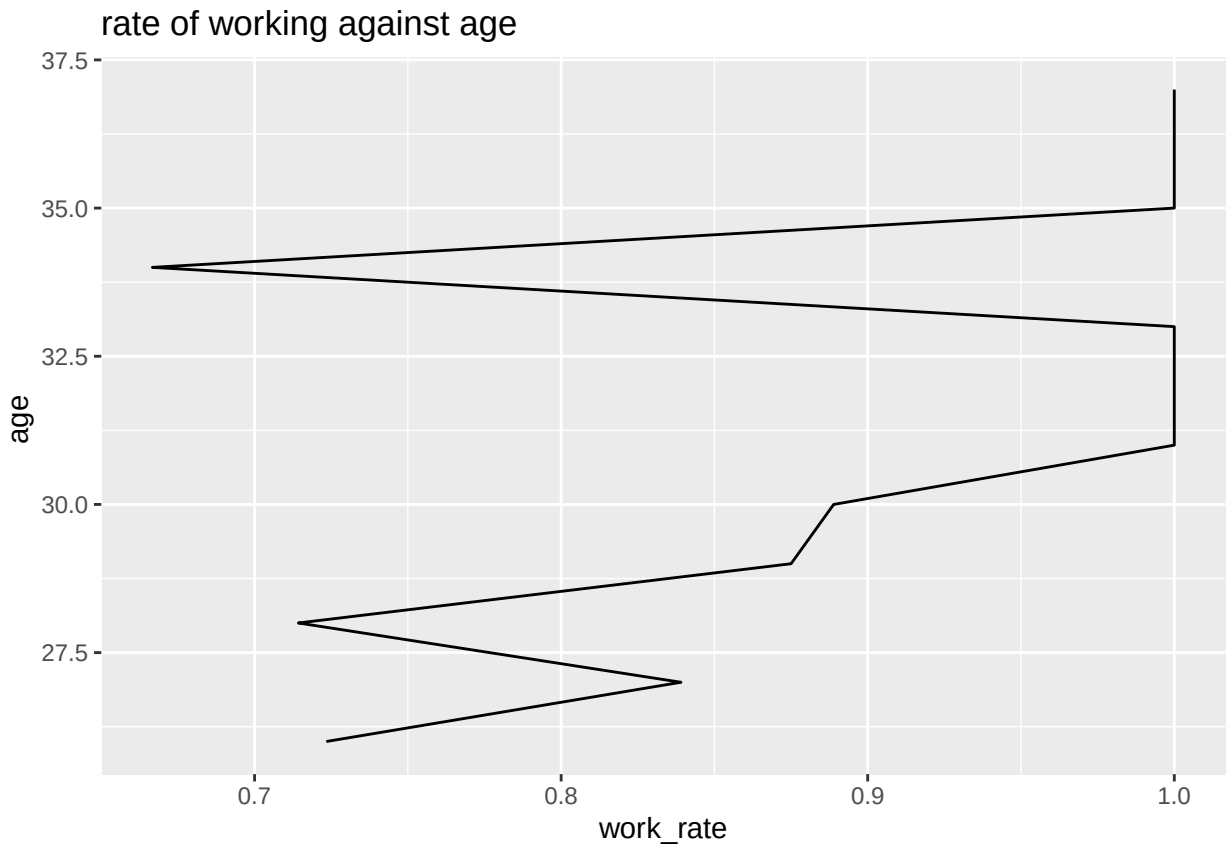
```
work_RSFD2016 <- PSFD2016 %>% group_by(age) %>% summarise(work_rate = mean(work))
```

```
library("ggplot2")
```

```
## Warning:   'ggplot2'   R   4.2.2
```

```
work_age <- ggplot(work_RSFD2016) +
#geom_bar(mapping = aes(y=work_rate, x=age), stat = "identity",width = 0.5)+
geom_line(mapping = aes(y=work_rate, x=age)) +
labs(title="rate of working against age")+
coord_flip()+
ylab('work_rate')+
xlab('age ')


plot(work_age)
```




```
rm(list = ls())
```

## 4.2.1: set value

```
mu0 <- 10
mu1 <- 9
```

```r
var0 <- 1
var1 <- 2
var01 <- 0.6
c <- 1
```

### 4.2.2: create eplison and store in data.table

```r
library('dplyr')
library('data.table')
```

```
## Warning:   'data.table'   R   4.2.2
```

```
##
##    'data.table'
```

```
##       'package:dplyr':
##
##     between, first, last
```

```r
simulation <- data.table(
  eplison0 = rnorm(10^7, 0, var0),
  eplison1 = rnorm(10^7, 0, var1)
)
```

```r
head(simulation)
```

```
##      eplison0   eplison1
## 1: -0.8861022 -1.3282397
## 2:  1.9921587  0.3519611
## 3:  0.6357725 -1.6665748
## 4: -0.4005967  0.8721765
## 5: -0.1781571 -2.0013408
## 6: -2.1244524 -1.7359779
```

### 4.2.3: create W0 and W1

```r
# W0 = u0 + eplison0
# w1 = u1 + eplison1

simulation[, w0:= mu0+eplison0]
simulation[, w1:= mu1+eplison1]
```

```r
head(simulation)
```

```
##      eplison0   eplison1        w0       w1
## 1: -0.8861022 -1.3282397  9.113898 7.671760
## 2:  1.9921587  0.3519611 11.992159 9.351961
## 3:  0.6357725 -1.6665748 10.635773 7.333425
## 4: -0.4005967  0.8721765  9.599403 9.872177
## 5: -0.1781571 -2.0013408  9.821843 6.998659
## 6: -2.1244524 -1.7359779  7.875548 7.264022
```

### 4.2.4: create I

```
# Migrate(I=1) if w1>(w0+c)

simulation[w1-w0-c>0, I:= 1]
simulation[w1-w0-c<=0, I:= 0]
```

```
head(simulation)
```

```
##      eplison0   eplison1        w0       w1 I
## 1: -0.8861022 -1.3282397  9.113898 7.671760 0
## 2:  1.9921587  0.3519611 11.992159 9.351961 0
## 3:  0.6357725 -1.6665748 10.635773 7.333425 0
## 4: -0.4005967  0.8721765  9.599403 9.872177 0
## 5: -0.1781571 -2.0013408  9.821843 6.998659 0
## 6: -2.1244524 -1.7359779  7.875548 7.264022 0
```

### 4.2.5: create conditional mean by data

```
conditionmean_bydata<- simulation[,
                                  .(condi_w0_bydata = mean(w0), #condi_w0_bydata is E(w0|I) cal by data
                                    condi_w1_bydata = mean(w1), #condi_w1_bydata is E(w1|I) cal by data
                                    Q0_bydata = mean(eplison0), #Q0_bydata is E(eplison0|I) cal by data
                                    Q1_bydata = mean(eplison1)),#Q1_bydata is E(eplison1|I) cal by data
                                  by = I]
```

```
conditionmean_bydata
```

```
##    I condi_w0_bydata condi_w1_bydata  Q0_bydata  Q1_bydata
## 1: 0       10.146811        8.413098  0.1468107 -0.5869023
## 2: 1        9.355504       11.579110 -0.6444964  2.5791103
```

```
# The first row(I=1) is what we want to know.
```

### 4.2.6: create conditional mean by RHS

```r
# get z by data
simulation[, v := eplison1-eplison0]
val_v <- simulation[, var(v)]
z <- (mu0-mu1+c)/val_v

# where we know E(w1|I) = mu1 + var1*E((eplison1/var1)|(v/var_v > z))
# and E(w0|I) = mu0 + var0*E((eplison0/var0)|(v/var_v > z))
simulation[v/var(v) > z, condi:= 1]
simulation[v/var(v) <= z, condi:= 0]
```

```r
E1 <- simulation[,mu1+ var1*mean(eplison1/var1), by = condi]

conditionmean_bydRHD<- simulation[,
# condi_w0_byRHS is E(w0|I) cal by E(w0|I) = mu0 + var0*E((eplison0/var0)|(v/var_v > z))
                           .(condi_w0_byRHS = mu0+ var0*mean(eplison0/var0),
# condi_w1_byRHS is E(w1|I) cal by E(w1|I) = mu1 + var1*E((eplison1/var1)|(v/var_v > z))
                             condi_w1_byRHS = mu1+ var1*mean(eplison1/var1)),
                             by = condi]

# Q0_byRHS is E(eplison0|I) cal by calculation(E(w0|I)-mu0)
conditionmean_bydRHD[,Q0_byRHS := (condi_w0_byRHS - mu0)]
# Q1_byRHS is E(eplison1|I) cal by calculation(E(w1|I)-mu1)
conditionmean_bydRHD[,Q1_byRHS := (condi_w1_byRHS - mu1)]
```

```r
conditionmean_bydRHD
```

```
##    condi condi_w0_byRHS condi_w1_byRHS   Q0_byRHS   Q1_byRHS
## 1:     0      10.146811       8.413098  0.1468107 -0.5869023
## 2:     1       9.355504      11.579110 -0.6444964  2.5791103
```

## 4.2.7

```r
# E(w1,I=1) & E(w0,I=0) is observed in real world,
# E(w1,I=0) & E(w0,I=1) is not observed in real world,
```