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
```

```
PSFD2020 <- read_sav("C:/Users/user/Downloads/C00377_1/psfd_rr2020_v202208_spss.sav")
head(PSFD2020)
```

```
## # A tibble: 6 x 1,063
##     x01 x01a      x01b    x01c     x01d  x01e x01f      x02 x03z01 x03z02 x03z03
##   <dbl> <dbl+lbl> <dbl+l> <dbl+l> <dbl> <dbl> <dbl+l> <dbl>  <dbl>  <dbl>  <dbl>
## 1 10020 1 [1  ~ 1 [1 1~ 1 [1 ~  1002     0 4 [4 ~  2020      1     26     23
## 2 10021 3 [3  ~ 6 [6 ~ 1 [1 ~  1002     1 4 [4 ~  2020      1     18     23
## 3 10022 3 [3  ~ 6 [6 ~ 1 [1 ~  1002     2 4 [4 ~  2020      1      9      8
## 4 10023 3 [3  ~ 6 [6 ~ 1 [1 ~  1002     3 4 [4 ~  2020      1     10      6
## 5 10040 1 [1  ~ 1 [1 1~ 1 [1 ~  1004     0 2 [2 ~  2020      2      7     20
## 6 10041 3 [3  ~ 6 [6 ~ 1 [1 ~  1004     1 4 [4 ~  2020      1     15      3
## # ... with 1,052 more variables: x03z04 <dbl>, x05 <dbl+lbl>, a01 <dbl+lbl>,
## #   a02a <dbl+lbl>, a02b <dbl+lbl>, a03a <dbl+lbl>, a03b <dbl+lbl>,
## #   a03c <dbl+lbl>, a03d <dbl+lbl>, w01a <dbl+lbl>, w01b <dbl+lbl>,
## #   w01c <dbl+lbl>, w01d <dbl+lbl>, w01e <dbl+lbl>, w01f <dbl+lbl>,
## #   w01g <dbl+lbl>, w02a <dbl+lbl>, w02b <dbl+lbl>, a05b01 <dbl+lbl>,
## #   a05b02 <dbl+lbl>, a05b03 <dbl+lbl>, a05b04 <dbl+lbl>, w03 <dbl+lbl>,
## #   w04 <dbl+lbl>, w05a <dbl+lbl>, w05b <dbl+lbl>, w05c01 <dbl+lbl>, ...
```

```
#check if   (x02) == 2020
PSFD2020[PSFD2020$x02 != 2020,]
```

```
## # A tibble: 0 x 1,063
## # ... with 1,063 variables: x01 <dbl>, x01a <dbl+lbl>, x01b <dbl+lbl>,
## #   x01c <dbl+lbl>, x01d <dbl>, x01e <dbl>, x01f <dbl+lbl>, x02 <dbl>,
## #   x03z01 <dbl>, x03z02 <dbl>, x03z03 <dbl>, x03z04 <dbl>, x05 <dbl+lbl>,
## #   a01 <dbl+lbl>, a02a <dbl+lbl>, a02b <dbl+lbl>, a03a <dbl+lbl>,
## #   a03b <dbl+lbl>, a03c <dbl+lbl>, a03d <dbl+lbl>, w01a <dbl+lbl>,
## #   w01b <dbl+lbl>, w01c <dbl+lbl>, w01d <dbl+lbl>, w01e <dbl+lbl>,
## #   w01f <dbl+lbl>, w01g <dbl+lbl>, w02a <dbl+lbl>, w02b <dbl+lbl>, ...
```

```
#calculate age
##a02a:
##  =2020 =  109 )

PSFD2020$age <- 109 - PSFD2020$a02a + 1

#w03:     ? 1:  2 3:   4:
##only consider work at market(1).
PSFD2020$work <- as.integer(PSFD2020$w03 == 1)
```

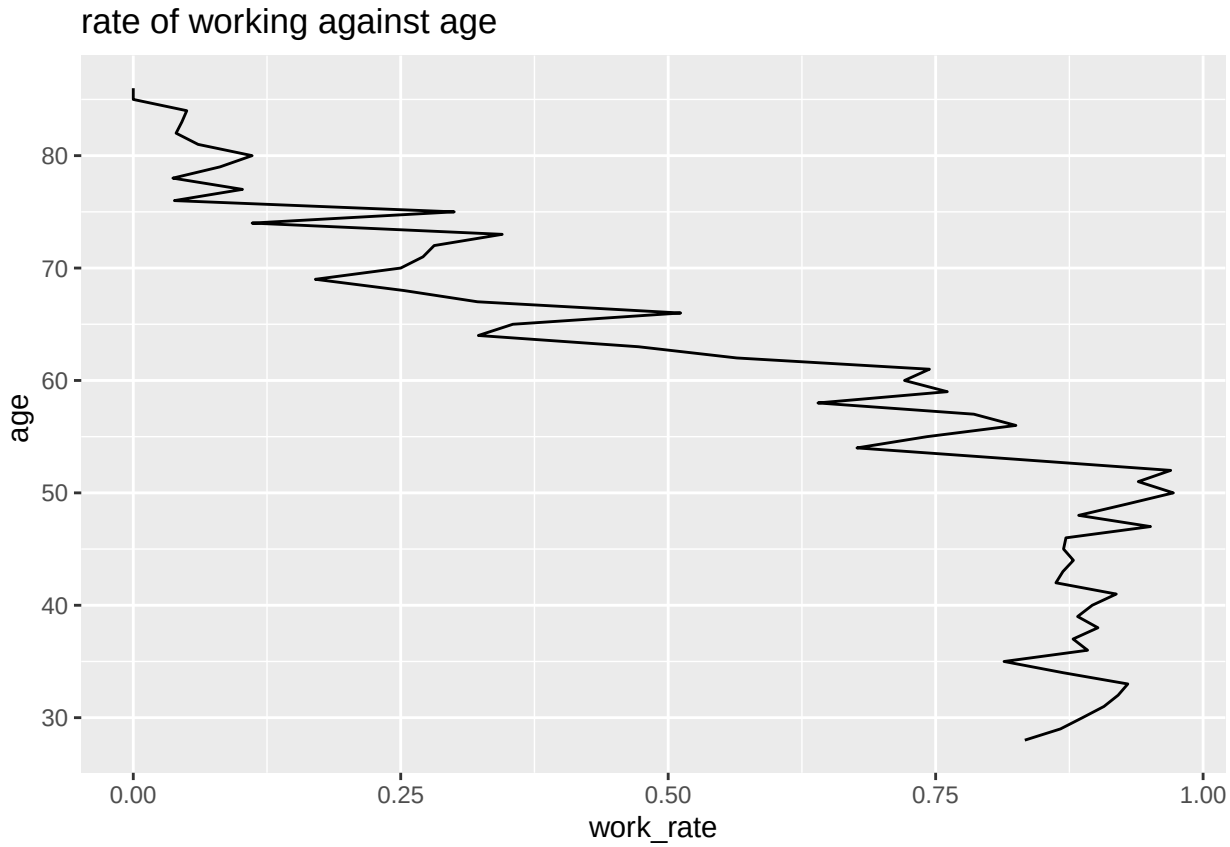
```
work_RSFD2020 <- PSFD2020 %>% group_by(age) %>% summarise(work_rate = mean(work))
```

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

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

```r
work_age <- ggplot(work_RSFD2020) +
#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)
```




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

4.2.1: set value

```r
mu0 <- 10
mu1 <- 9
var0 <- 1
var1 <- 2
var01 <- 0.6
c <- 1
```

4.2.2: create eplison and store in data.table

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

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

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

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

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

```
head(simulation)
```

```
##       eplison0   eplison1
## 1:  0.93648312 -0.7542338
## 2:  1.07446506  4.3505676
## 3:  0.45959078  0.7415615
## 4:  1.13750641 -5.7909969
## 5: -1.45201266  1.6141097
## 6:  0.07024064  4.1294119
```

4.2.3: create W0 and W1

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

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

```
head(simulation)
```

```
##       eplison0   eplison1        w0        w1
## 1:  0.93648312 -0.7542338 10.936483  8.245766
## 2:  1.07446506  4.3505676 11.074465 13.350568
## 3:  0.45959078  0.7415615 10.459591  9.741561
## 4:  1.13750641 -5.7909969 11.137506  3.209003
## 5: -1.45201266  1.6141097  8.547987 10.614110
## 6:  0.07024064  4.1294119 10.070241 13.129412
```

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:  0.93648312 -0.7542338 10.936483  8.245766 0
## 2:  1.07446506  4.3505676 11.074465 13.350568 1
## 3:  0.45959078  0.7415615 10.459591  9.741561 0
## 4:  1.13750641 -5.7909969 11.137506  3.209003 0
## 5: -1.45201266  1.6141097  8.547987 10.614110 1
## 6:  0.07024064  4.1294119 10.070241 13.129412 1
```

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       10.146603        8.413288  0.1466030 -0.5867123
## 2: 1        9.356412       11.578837 -0.6435878  2.5788366
```

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

4.2.6: create conditional mean by RHS

```
# 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      10.146603       8.413288  0.1466030 -0.5867123
## 2:     1       9.356412      11.578837 -0.6435878  2.5788366
```

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,
```