```
Raking
黃楷捷
2023-04-26
 # 多變項反覆加權 #
 library(car)
 library(XML)
 library(haven)
 library(tidyr)
 library(expss)
 library(survey)
 library(weights)
 library(stringr)
 library(data.table)
 ############
 # 調查資料 #
 ############
 # 匯入調查資料 #
 setwd("/Users/kyle/Desktop/NCCU/111-2/三1234社會研究方法/HW")
 tscs201 <- read_dta('tscs201.dta') |> setDT()
 tscs201 \leftarrow tscs201[, Sex := factor(x = v1, levels = c(1:2), labels = c("男", "女"))]
 tscs201 <- apply labels(data = tscs201, Sex = "性別")
 # 重新編碼變項 #
 tscs201 \leftarrow (copy(tscs201)[v2y < 100, Age := 109 - v2y]
                         [v2y > 100, Age := grep("[0-9]", v2r)]
                         [, Age := Recode(Age, recodes = '18:29="18-29歳"; 30:39="30-39歳";
                                                         40:49="40-49歲"; 50:59="50-59歲";
                                                         60:69="60-69歲"; else="70歲以上"')]
                         [v118 != 97, ])
 tscs201 <- apply labels(data = tscs201, Age = "年齡")
 Sample <- tscs201[, .(Sex, Age, License = v118, TW = v41, id = c(1:nrow(tscs201)))]
 ############
 # 母體資料 #
 ############
 # 匯入人口母體資料 #
 setwd('/Users/kyle/Desktop/NCCU/社會調查師/R-Data-Processing/DataSets')
 pop <- read.csv("10912村里戶數、單一年齡人口.csv", skip = 1)
 names(pop) <- str_replace_all(string = names(pop), pattern = c("X" = "", "\\." = ""))</pre>
 pop <- subset(x = pop, select = -c(統計年月, 戶數, 人口數, 人口數男, 人口數女))
 popByAge <- pivot_longer(data = pop,</pre>
                         cols = str_subset(string = names(pop), pattern = "^[0-9]"),
                         names to = "Group", values to = "Population") |> setDT()
 popByAge <- (popByAge[, .(Area = str_sub(區域別, 1, 3),
                          Sex = str_extract(string = Group, pattern = "[男女]$"),
                          年龄 = as.numeric(str_extract(string = Group, pattern = "(^[0-9]{1,3})")),
                          人口數 = as.numeric(Population))]
                     [年龄 >= 18 , Age := Recode(年龄, recodes = '18:29="18-29歳"; 30:39="30-39歳";
                                                                40:49="40-49歳"; 50:59="50-59歳";
                                                                 60:69="60-69歲"; else="70歲以上"')])
 # 計算母體性別及年齡比例 #
 Population <- popByAge[年齡 >= 18, .(人口數 = sum(人口數)), by = .(Sex, Age)]
 sex <- Population[, .(人口數比例 = 人口數 / sum(人口數), Sex)][, .(Freq = sum(人口數比例)), by = .(Sex)] |> t()
 age <- Population[, .(人口數比例 = 人口數 / sum(人口數), Age)][, .(Freq = sum(人口數比例)), by = .(Age)] |> t()
 # 使用卡方檢定檢查分佈情形 #
 chisq.test(table(tscs201$Sex), p = as.numeric(sex[2, ]))
 ##
    Chi-squared test for given probabilities
 ## data: table(tscs201$Sex)
 ## X-squared = 9.4976, df = 1, p-value = 0.002057
 chisq.test(table(tscs201$Age), p = round(as.numeric(age[2, ]), 2))
    Chi-squared test for given probabilities
 ## data: table(tscs201$Age)
 ## X-squared = 46.908, df = 5, p-value = 5.933e-09
 # 依照母體比例計算期望值 #
 sex2 <- Population[, .(人口數比例 = 人口數 / sum(人口數), Sex)][, .(Freq = nrow(Sample) * sum(人口數比例)), by = .(Se
 age2 <- Population[, .(人口數比例 = 人口數 / sum(人口數), Age)][, .(Freq = nrow(Sample) * sum(人口數比例)), by = .(Ag
 e)]
 ############
 # 反覆加權 #
 ############
 ## 手動進行反覆加權 ##
 # 加權性別 #
 sex3 <- prop.table(table(Sample$Sex)) |> data.frame() |> t()
 sexW <- data.frame(Sex = c('男', '女'), WS = as.numeric(sex[2,]) / as.numeric(sex3[2,]))
 Sample <- merge(Sample, sexW, by = 'Sex')</pre>
 Sample[, Sex := factor(Sex, levels = c('男', '女'))]
 # 檢查加權性別後分佈 #
 chisq.test(wpct(Sample$Sex, Sample$WS) * nrow(Sample), p = as.numeric(sex[2, ]))
    Chi-squared test for given probabilities
 ## data: wpct(Sample$Sex, Sample$WS) * nrow(Sample)
 \#\# X-squared = 7.2215e-12, df = 1, p-value = 1
 chisq.test(wpct(Sample$Age, Sample$WS) * nrow(Sample), p = round(as.numeric(age[2, ]), 2))
    Chi-squared test for given probabilities
 ## data: wpct(Sample$Age, Sample$WS) * nrow(Sample)
 ## X-squared = 43.445, df = 5, p-value = 3.002e-08
 # 加權年齡 #
 SexAgeCount <- Sample[,.(.N), by = .(Age, Sex)]
 AgeProb <- merge(SexAgeCount, sexW)</pre>
 AgeProb \leftarrow AgeProb[, N := N * WS][, \cdot(N = sum(N)), by = \cdot(Age)][, ProbA := N / sum(N)][]
 A <- Population[, .(人口數比例 = 人口數 / sum(人口數), Age)][, .(FreqA = sum(人口數比例)), by = .(Age)]
 ageW <- merge(AgeProb, A)</pre>
 ageW <- ageW[, .(Age, WA = FreqA / ProbA)]</pre>
```

```
Sample <- merge(Sample, ageW)</pre>
# 檢查加權年齡後分佈 #
chisq.test(wpct(Sample$Sex, Sample$WA) * nrow(Sample), p = as.numeric(sex[2, ]))
```

```
Chi-squared test for given probabilities
## data: wpct(Sample$Sex, Sample$WA) * nrow(Sample)
## X-squared = 5.9532, df = 1, p-value = 0.01469
```

chisq.test(wpct(Sample\$Age, Sample\$WA) * nrow(Sample), p = round(as.numeric(age[2,]), 2))

```
##
   Chi-squared test for given probabilities
## data: wpct(Sample$Age, Sample$WA) * nrow(Sample)
\#\# X-squared = 0.41704, df = 5, p-value = 0.9948
```

```
# 計算總權數 #
Sample[, weights2 := WS * WA]
# 檢查總權數加權後分佈 #
chisq.test(wpct(Sample$Sex, Sample$weights2) * nrow(Sample), p = as.numeric(sex[2, ]))
##
   Chi-squared test for given probabilities
```

```
## data: wpct(Sample$Sex, Sample$weights2) * nrow(Sample)
## X-squared = 0.41754, df = 1, p-value = 0.5182
```

chisq.test(wpct(Sample\$Age, Sample\$weights2) * nrow(Sample), p = round(as.numeric(age[2,]), 2))

4 97

2

3 ## 118.4 805.6 641.1 194.5 92.1 2.3

```
Chi-squared test for given probabilities
## data: wpct(Sample$Age, Sample$weights2) * nrow(Sample)
## X-squared = 0.49947, df = 5, p-value = 0.9921
```

```
##########################
# v41台灣社會發展 #
######################
table(Sample$TW)
```

```
1 2 3 4 97 98
## 118 784 649 197 104 2
```

```
(wpct(Sample$TW, Sample$weights) * nrow(Sample)) |> round(1)
```