Simulation of data

YK

2020/10/8

Data Simulation

We should have total 6 variables obtained from the results of survey.

Age

(1) For age, from Statistics Canada we've known the number of people of each age group in Ontario population, in this case we divide population into 4 groups(0-18,19-49,50-89,90 and over), and by calculation the percentages are 21%, 41.1%, 37% and 0.9% respectively.

```
# Size of population(Number of simulations)
N<-100000
# Age
# We assume in each age group, ages appears at equal probability, so it follows a uniform distribution.
Age_0_18<-runif(N* 0.21, min = 0, max = 18)
Age_19_49<-runif(N*0.411, min = 19, max = 49)
Age_50_89<-runif(N*0.37, min = 50, max = 89)
Age_90_over<-runif(N*0.009, min = 90, max = 110)
#This is the simulated Age data
Age_all<-c(Age_0_18, Age_19_49, Age_50_89, Age_90_over)
# Shuffle the data
ShAge_all<-sample(Age_all)
```

#Gender

(2)For Gender, we assume the ratio of female and male is 1:1.

```
#Gender
gender<-c("Female","Male")
Gender_all<-sample(gender,N,replace = T)
```

#District

(3) For district the person is living in, there are total 124 districts in Ontario, we assume each person has the equal probability to live in any district.

```
#District
District<-sample(1:124, N, replace = T)
```

#Supporting Political party

(4) We referenced the results of election in 2019 and will simulate according to the vote share in Ontario.

*LIB: 41.4%

*CON: 33.2%

*NDP: 16.8%

*GRN: 6.2%

*IND: 0.4%

*PP: 1.6%

*OTH: 0.4%

```
#Political Party supported
#Simulate according to the past vote share
Poli_Parties<-c("LIB","CON","NDP","GRN","IND","PP","OTH")
PoliticalP_supported<-sample(Poli_Parties,N,prob = c(0.414, 0.332, 0.168, 0.062, 0.004, 0.016, 0.004), r
```

Education Level

(5)We referenced from the 2016 education attainment status in Canada, and will simulate according to this.

*Bachelor's degree or higher: 28.5%

*University below bachelor's 3.1%

*College diploma 22.4%

*Apprenticeship or other trades certificate 10.8%

*High school diploma 23.7%

*No certificate, diploma or degree 11.5%

```
#Education level
Edu_levels<-c("Bachelor or higher","University below bachelor's", "College diploma","Apprenticeship or
Edu_level_all<-sample(Edu_levels, N, prob = c(0.285, 0.031, 0.224, 0.108, 0.237, 0.115),replace = T)
```

Income

(6)We simulate income in the same way we simulate age.