SNAP Penetration Analyses

Tzu-Hsuan (Jessica) Lin, Koller Adzick, Katherine Li, Allen Hu

2020/11/30

```
library(tidyverse)
library(dplyr)
library(readxl)
library(tidyr)
```

Read in data

Household size per tract

With data from 2019 American Community Survey, we extracted the size and margin error of household in each tract in 2018.

```
# tract household size

tract <- read_excel("hennepin_county_all_tracts_household_size_us_census_2018 - Copy.xlsx", skip=9)

tract <- tract[3, grepl("Tract", names(tract))]
tract <- gather(tract)
tract$key <- gsub(", Hennepin County, Minnesota", "", tract$key)
tract$key <- gsub("Census Tract ", "", tract$key)
tract$value <- gsub(",", "", tract$value)

tract$key <- as.numeric(tract$key)
tract$value <- as.numeric(tract$value)

tract <- tract %>%
  rename(tract_name = key,
         household_size = value)


# tract household margin
margin <- read_excel("hennepin_county_all_tracts_household_size_us_census_2018 - Copy.xlsx", skip=9)
margin <- margin[3,]
margin <- gather(margin)
margin <- margin[grepl("±",margin$value), ]

margin$value <- gsub("±", "", margin$value)

margin$key <- as.numeric(margin$key)
margin$value <- as.numeric(margin$value)
```

```
margin <- margin %>%
  rename(household_margin = value)


tract <- cbind(tract, margin$household_margin)
tract <- tract %>%
  rename(household_margin = `margin$household_margin`)
```

Income level by ethnicity

Again, using data from American Community Survey, we extracted ethnicity level income.

```
ethnicity <- read_excel("ACSST1Y2019.S1903-2020-11-25T211057.xlsx", skip=11)
ethnicity <- ethnicity[, c(1, 6, 7)]

ethnicity <- ethnicity %>%
  rename(median_income_ethnicity = `Estimate...6`,
         margin_income_ethnicity = `Margin of Error...7`)

ethnicity$Label <- gsub("(^[[:space:]]+|[[:space:]]+$)", "", ethnicity$Label)
ethnicity$margin_income_ethnicity <- gsub("±", "",
                                          ethnicity$margin_income_ethnicity)
ethnicity$median_income_ethnicity <- gsub(',', "",
                                          ethnicity$median_income_ethnicity)
ethnicity$margin_income_ethnicity <- gsub(',', "",
                                          ethnicity$margin_income_ethnicity)

ethnicity$median_income_ethnicity <- as.numeric(ethnicity$median_income_ethnicity)
ethnicity$margin_income_ethnicity <- as.numeric(ethnicity$margin_income_ethnicity)

ethnicity <- ethnicity[4:12, ]


# replace names
ethnicity$Label <- gsub("Black or African American",
                        "Black/African American",
                        ethnicity$Label)
ethnicity$Label <- gsub("American Indian and Alaska Native",
                        "American Indian or Alaskan Native",
                        ethnicity$Label)
ethnicity$Label <- gsub("Asian",
                        "Asian/Pacific Islander",
                        ethnicity$Label)
ethnicity$Label <- gsub("Some other race",
                        "Other or Unknown",
                        ethnicity$Label)
ethnicity$Label <- gsub("Hispanic or Latino origin (of any race)",
                        "Hispanic or Latino",
                        ethnicity$Label,
                        fixed=TRUE)
```

Income limit function

With clearer definition of how the income limit should be set (165% poverty line), we leveraged the linear model function and found out the exact limit for each tract's household. (eg, avg 1.7 people per household, then the monthly income limit is roughly 2186)

```
limit <- c(1755, 2371, 2987, 3603, 4219, 4835, 5451, 6067)
ppl <- c(1, 2 ,3, 4, 5, 6, 7, 8)

income <- data.frame(ppl, limit)

summary(lm(limit~ppl))
```

```
##
## Call:
## lm(formula = limit ~ ppl)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.027e-13 -1.870e-13  4.021e-14  1.782e-13  2.476e-13
##
## Coefficients:
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept) 1.139e+03  1.774e-13 6.421e+15   <2e-16 ***
## ppl         6.160e+02  3.513e-14 1.753e+16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.277e-13 on 6 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.075e+32 on 1 and 6 DF,  p-value: < 2.2e-16
```

```
# income limit function: limit= 1139 + 616 * ppl
```

Snap people count and census data

FFIEC dataset contains 2016-2019 estimations for population and tract level income.

```
snap <- read_excel('SNAP Summary data for LiveCase fall 2020.xlsx')
census <- read_csv('2010_Census_Tracts.csv')
```

Joining datasets for 2016

```
ffiec <- read_excel('2016 FFIEC Summary Census Demographic Information.xlsx')

df <- snap %>%
  full_join(census, by = c('tract' = 'GEOID10'), suffix = c('snap', 'census'))
df <- df %>%
  full_join(ffiec, by = c('NAME10' = 'Tract Code'), suffix = c('df', 'ffiec'))
```

```
df_2016 <- df %>%
  filter(elig_month > '2015-12-31', elig_month < '2017-01-01')
df_2016 <- df_2016 %>%
  filter(`Tract Income Level` != 'Unknown')

df_2016 <- df_2016 %>%
  select('elig_month', 'tract', 'NAME10', 'race_ethnicity', 'people', 'ALAND10',
         'TRACT', 'POP_TOTAL', 'POP1_WHT', 'POP1_BLK',
         'POP1_AMIND', 'POP1_ASN', 'POP1_HAWPA',  'POP1_OTHR',
         'POP_HSPLAT', 'Tract Income Level',
         '2016 Est. Tract Median Family Income', 'Tract Population') %>%
  rename('median_income_ffiec' = '2016 Est. Tract Median Family Income',
         'pop_ffiec' = 'Tract Population')


df_2016 <- df_2016 %>%
  left_join(tract, by=c('NAME10' = 'tract_name'))
df_2016 <- df_2016 %>%
  left_join(ethnicity, by=c('race_ethnicity' = 'Label'))


df_2016['ppl_per_household'] <- (df_2016['pop_ffiec'] / df_2016['household_size'])
df_2016['monthly_income_limit'] <- 1139 + 616 * df_2016['ppl_per_household']
df_2016['annual_income_limit'] <- df_2016['monthly_income_limit'] * 12
```

Joining datasets for 2017

```
ffiec <- read_excel('2017 FFIEC Summary Census Demographic Information.xlsx')

df <- snap %>%
  full_join(census, by = c('tract' = 'GEOID10'), suffix = c('snap', 'census'))
df <- df %>%
  full_join(ffiec, by = c('NAME10' = 'Tract Code'), suffix = c('df', 'ffiec'))

df_2017 <- df %>%
  filter(elig_month > '2016-12-31', elig_month < '2018-01-01')
df_2017 <- df_2017 %>%
  filter(`Tract Income Level` != 'Unknown')

df_2017 <- df_2017 %>%
  select('elig_month', 'tract', 'NAME10', 'race_ethnicity', 'people', 'ALAND10',
         'TRACT', 'POP_TOTAL', 'POP1_WHT', 'POP1_BLK',
         'POP1_AMIND', 'POP1_ASN', 'POP1_HAWPA',  'POP1_OTHR',
         'POP_HSPLAT', 'Tract Income Level',
         '2017 Est. Tract Median Family Income', 'Tract Population') %>%
  rename('median_income_ffiec' = '2017 Est. Tract Median Family Income',
         'pop_ffiec' = 'Tract Population')


df_2017 <- df_2017 %>%
  left_join(tract, by=c('NAME10' = 'tract_name'))
```

```
df_2017 <- df_2017 %>%
  left_join(ethnicity, by=c('race_ethnicity' = 'Label'))


df_2017['ppl_per_household'] <- (df_2017['pop_ffiec'] / df_2017['household_size'])
df_2017['monthly_income_limit'] <- 1139 + 616 * df_2017['ppl_per_household']
df_2017['annual_income_limit'] <- df_2017['monthly_income_limit'] * 12
```

Joining datasets for 2018

```
ffiec <- read_excel('2018 FFIEC Summary Census Demographic Information.xlsx')

df <- snap %>%
  full_join(census, by = c('tract' = 'GEOID10'), suffix = c('snap', 'census'))
df <- df %>%
  full_join(ffiec, by = c('NAME10' = 'Tract Code'), suffix = c('df', 'ffiec'))

df_2018 <- df %>%
  filter(elig_month > '2017-12-31', elig_month < '2019-01-01')
df_2018 <- df_2018 %>%
  filter(`Tract Income Level` != 'Unknown')

df_2018 <- df_2018 %>%
  select('elig_month', 'tract', 'NAME10', 'race_ethnicity', 'people', 'ALAND10',
         'TRACT', 'POP_TOTAL', 'POP1_WHT', 'POP1_BLK',
         'POP1_AMIND', 'POP1_ASN', 'POP1_HAWPA',  'POP1_OTHR',
         'POP_HSPLAT', 'Tract Income Level',
         '2018 Est. Tract Median Family Income', 'Tract Population') %>%
  rename('median_income_ffiec' = '2018 Est. Tract Median Family Income',
         'pop_ffiec' = 'Tract Population')


df_2018 <- df_2018 %>%
  left_join(tract, by=c('NAME10' = 'tract_name'))
df_2018 <- df_2018 %>%
  left_join(ethnicity, by=c('race_ethnicity' = 'Label'))


df_2018['ppl_per_household'] <- (df_2018['pop_ffiec'] / df_2018['household_size'])
df_2018['monthly_income_limit'] <- 1139 + 616 * df_2018['ppl_per_household']
df_2018['annual_income_limit'] <- df_2018['monthly_income_limit'] * 12
```

Joining datasets for 2019

```
ffiec <- read_excel('2019 FFIEC Summary Census Demographic Information.xlsx')

df <- snap %>%
  full_join(census, by = c('tract' = 'GEOID10'), suffix = c('snap', 'census'))
df <- df %>%
```

```
  full_join(ffiec, by = c('NAME10' = 'Tract Code'), suffix = c('df', 'ffiec'))

df_2019 <- df %>%
  filter(elig_month > '2018-12-31', elig_month < '2020-01-01')
df_2019 <- df_2019 %>%
  filter(`Tract Income Level` != 'Unknown')

df_2019 <- df_2019 %>%
  select('elig_month', 'tract', 'NAME10', 'race_ethnicity', 'people', 'ALAND10',
         'TRACT', 'POP_TOTAL', 'POP1_WHT', 'POP1_BLK',
         'POP1_AMIND', 'POP1_ASN', 'POP1_HAWPA',  'POP1_OTHR',
         'POP_HSPLAT', 'Tract Income Level',
         '2019 Est. Tract Median Family Income', 'Tract Population') %>%
  rename('median_income_ffiec' = '2019 Est. Tract Median Family Income',
         'pop_ffiec' = 'Tract Population')


df_2019<- df_2019 %>%
  left_join(tract, by=c('NAME10' = 'tract_name'))
df_2019 <- df_2019 %>%
  left_join(ethnicity, by=c('race_ethnicity' = 'Label'))


df_2019['ppl_per_household'] <- (df_2019['pop_ffiec'] / df_2019['household_size'])
df_2019['monthly_income_limit'] <- 1139 + 616 * df_2019['ppl_per_household']
df_2019['annual_income_limit'] <- df_2019['monthly_income_limit'] * 12
```

Join datasets for all years

```
df <- rbind(df_2016, df_2017, df_2018, df_2019)
```

Penetration rate calculation

Adjust tract income based on ethnicity income ratio

The idea is to use a ratio, based on ethnicity. `ethnicity_avg` is the average of 6 ethnicities (that appeared in the snap dataset). Then, ethnicity income / average income for all ethnicity would give us a ratio (eg. white people are earning 90321 / 63151.33 = 1.43 more than average). Then, for each tract, use their tract income to multiply its ratio. Thus give us an adjusted income. Then apply log-normal disctibution to calculate "how many people are under its corresponding income boundary".

```
# calculate average for all ethnicity
ethnicity_avg <- ethnicity %>%
  filter((Label == 'White') | (Label == 'Black/African American') |
           (Label == 'American Indian or Alaskan Native') |
           (Label == 'Asian/Pacific Islander') |
           (Label == 'Other or Unknown') | (Label == 'Hispanic or Latino')) %>%
  summarize(avg = mean(median_income_ethnicity),
            std = sd(margin_income_ethnicity))
```

```r
df <- cbind(df, ethnicity_avg)
df['adjusted_income'] <- (df['median_income_ffiec'] *
                                df['median_income_ethnicity'] /
                                df['avg'])


# calculate population under corresponding income boundary
pop_perc_by_adjusted_income <- c()

for (i in 1:nrow(df)){

  income <- as.numeric(df[i,'adjusted_income'])
  limit <- as.numeric(df[i,'monthly_income_limit'])
  error <- as.numeric(df[i,'margin_income_ethnicity'])

  perc_single <- plnorm(limit, meanlog = log(income),
                        sdlog = sqrt(log((error*error/income)^2 + 1)))
  # print(perc_single)
  pop_perc_by_adjusted_income = c(pop_perc_by_adjusted_income, perc_single)

}

df <- df %>%
  cbind(pop_perc_by_adjusted_income)


# calculate eligible people and penetration rate
df['pop_limit_by_adjusted_income'] <- round(df['pop_ffiec'] *
                                      df['pop_perc_by_adjusted_income'],0)
df['snap_perc_by_adjusted_income'] <- round(df['people']/
                                     df['pop_limit_by_adjusted_income'], 4)
df['snap_perc_by_adjusted_income'] <- (df['snap_perc_by_adjusted_income'] /
                                           df['ppl_per_household'])
```