Maternal health seeking behavior data_clean

Fuyu

Maternal health seeking behavior

Pregnancy and childbirth are natural and often eventful processes many women are at risk for developing complications during pregnancy and childbirth. Complications of pregnancy and childbirth are the leading causes of disability and death among women in the reproductive age (15-49) years in developing countries. Maternal health seeking behavior normally includes three aspects according to articles: 1) antenatal care(ANC). 2) birth of place 3) post-natal care(PNC). According to literatures I have read, most researches focus on antenatal care and delivery birth place. Only one research in Eastern Ethiopia studied PNC utilization.

Besides, most researches treat these three aspects separately. I will present dependent variables they used below:
1) ANC: utilize ANC or not (binary varibale); Number of visits. WHO recommended woman with pregancy visits ANC at least 4 times.
2) Delivery place: In healthy facilities or in home; Assited by professional health care workers or traditional care workers?
3) PNC: utilize ANC or not (binary varibale)

Logically, these three aspects complies a complete pregnancy process, and they are not determined independently. However, if we use regress later stage on previous stage, e.g. depentent variable is delivery place, independent variables are ANC utilization and other demographic, social-economic factors. Linearity and other statistical association between demographic, social-economic factors and ANC utilization will make the variation of estimation in Delivery place model too large to accpet.

Therefore, I try to do a clustering analysis on ANC, delivery place and PNC variables. I want to treat them as a whole and do a multinomial regression to explore which factors influence their choice pattern.

Read and Merge data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(foreign)
library(haven)
library(data.table)
```

```
##
```

```
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
hh<-read_dta("D:\\tang_data\\hh.dta")
hh<-dplyr::select(hh,HH1,HH2,HC1A,HC1B,HC2,HH7,helevel)
hh[]<-lapply(hh, unclass)
hh$HH<-paste(hh$HH1,hh$HH2,sep = "_")
sum(duplicated(hh$HH))
```

```
## [1] 0
```

```
wm<-read_dta("D:\\tang_data\\wm.dta")
wm<-dplyr::select(wm,LN,HH1,CM11,HH2,MA1,WM6Y,WB5,WB6A,WB6B,
                  welevel,insurance,wscoreu,wscorer,CM11,WB18,WM14,WM15,WB3Y,WB4,WDOBLC,MN2,MN5
                  ,MN20,PN5,PN17,MN19A,MN19C,MN19D,MN19F,MN19G,MN19H,MN19X,MN19Y,PN9,PN20,PN25A,PN25B,PN
```

Create Personal ID

```
wm$HH<-paste(wm$HH1,wm$HH2,sep = "_")
wm$HHL<-paste(wm$HH,wm$LN,sep = "_")
sum(duplicated(wm$HHL))
```

```
## [1] 0
```

```
wm[] <- lapply(wm, unclass)
sum(duplicated(wm$HH))
```

```
## [1] 5492
```

```
wm_new<-base::merge(wm,hh,by=c("HH","HH1","HH2"),all=F)
rm(hh,wm)
sum(duplicated(wm_new$HHL))
```

```
## [1] 0
```

We choose those who gave birth in last two years before the survey year

```
wm_new$year=floor(wm_new$WDOBLC/12)+1900
wm_new<-as.data.table(wm_new)
wm_use<-wm_new[year-WM6Y>=-2]
```

WM6Y: Survey Year; year: Birth year

```
table(wm_use$WM6Y,wm_use$year)
```

```
##
##        2015 2016 2017 2018
##   2017   97  120  129    8
##   2018    0 3242 4426 1872
```

```
rm(wm_new)
table(wm_use$CM17)
```

```
##
```

```
##    0    1
## 1306 8560
```

```
table(wm_use$WM6Y,wm_use$year)
```

```
##
##        2015 2016 2017 2018
##   2017   97  120  129    8
##   2018    0 3242 4426 1872
```

```
wm_use<-wm_use[!(WM6Y==2017&year==2018),]
table(wm_use$WM6Y,wm_use$year)
```

```
##
##        2015 2016 2017 2018
##   2017   97  120  129    0
##   2018    0 3242 4426 1872
```

clean data

MN2: attend ANC or not MN20: Delivery place

```
wm_clean<-wm_use[!is.na(MN2)&!is.na(MN20)]
wm_clean<-wm_use[MN2!=9&MN20!=9]
table(wm_clean$WM6Y,wm_clean$year)
```

```
##
##        2016 2017 2018
##   2017  119  129    0
##   2018 2045 4404 1850
```

```
rm(wm_use)
```

ac4:ANC at least 4 times

```
wm_clean$MN5[wm_clean$MN2==2]=0
table(wm_clean$MN5)#you could see there are some outliers
```

```
##
##    0    1    2    3    4    5    6    7    8    9   10   11   12   13   15   16
## 1935  372 1035 1981 1579  828  456  130   77   39   27    1   16    2    3   11
##   18   20   21   22   27   30   36   46   60   72   82   98   99
##    1    1    1    1    1    1    3    1    1    3    1   38    2
```

```
#drop non-response
wm_clean<-wm_clean[MN5!=98]
table(wm_clean$MN5)
```

```
##
##    0    1    2    3    4    5    6    7    8    9   10   11   12   13   15   16
## 1935  372 1035 1981 1579  828  456  130   77   39   27    1   16    2    3   11
##   18   20   21   22   27   30   36   46   60   72   82   99
##    1    1    1    1    1    1    3    1    1    3    1    2
```

```
# create new variable to see if the women vistis antenatal care at leats 4 times
wm_clean$ac4<-0
wm_clean$ac4[wm_clean$MN5>=4]<-1
```

```
# where give birth to
table(wm_clean$MN20)
```

```
##
##   11   12   21   22   23   26   31   32   36   96   99
## 2129  429  959 2099 1453   83  526  300  442   84    5
```

```
wm_clean<-wm_clean[MN20!=99]
```

probithpl: Delivery place

```
#create where probithpl to indicate the birth place
wm_clean$probithpl<-1
wm_clean$probithpl[wm_clean$MN20<=12|wm_clean$MN20==96]<-0
wm_clean$probithpl[wm_clean$PN7]
```

```
## numeric(0)
```

```
table(wm_clean$probithpl)
```

```
##
##    0    1
## 2642 5862
```

PNC utilization

```
#postnatal check
wm_clean<-wm_clean[PN25A!=9&PN25A!=8]
wm_clean<-wm_clean[PN25B!=9&PN25B!=8]
wm_clean<-wm_clean[PN25C!=9&PN25C!=8]
summary(wm_clean)
```

```
##       HH                 HH1              HH2              LN
##  Length:8467        Min.   :  1.0   Min.   : 1.00   Min.   : 1.000
##  Class :character   1st Qu.:205.0   1st Qu.: 8.00   1st Qu.: 2.000
##  Mode  :character   Median :396.0   Median :15.00   Median : 2.000
##                     Mean   :382.5   Mean   :15.16   Mean   : 2.164
##                     3rd Qu.:563.0   3rd Qu.:23.00   3rd Qu.: 2.000
##                     Max.   :721.0   Max.   :30.00   Max.   :24.000
##
##       CM11             MA1             WM6Y           WB5
##  Min.   : 1.000   Min.   :1.000   Min.   :2017   Min.   :1.000
##  1st Qu.: 2.000   1st Qu.:1.000   1st Qu.:2018   1st Qu.:1.000
##  Median : 3.000   Median :1.000   Median :2018   Median :1.000
##  Mean   : 3.863   Mean   :1.398   Mean   :2018   Mean   :1.219
##  3rd Qu.: 5.000   3rd Qu.:2.000   3rd Qu.:2018   3rd Qu.:1.000
##  Max.   :16.000   Max.   :9.000   Max.   :2018   Max.   :9.000
##
##       WB6A            WB6B          welevel        insurance    wscoreu
##  Min.   : 0.00   Min.   :1.000   Min.   :0.000   Min.   :1   Min.   :-1.616
##  1st Qu.:10.00   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2   1st Qu.:-0.865
##  Median :10.00   Median :3.000   Median :1.000   Median :2   Median :-0.289
##  Mean   :18.51   Mean   :3.276   Mean   :1.418   Mean   :2   Mean   :-0.057
##  3rd Qu.:31.00   3rd Qu.:5.000   3rd Qu.:2.000   3rd Qu.:2   3rd Qu.: 0.683
##  Max.   :40.00   Max.   :6.000   Max.   :4.000   Max.   :9   Max.   : 2.868
```

```
##  NA's   :1849    NA's   :1851                                NA's   :6232
##     wscorer             WB18        WM14            WM15            WB3Y
##  Min.   :-1.4233   Min.   :1   Min.   :1.000   Min.   :1.000   Min.   :1968
##  1st Qu.:-0.6032   1st Qu.:2   1st Qu.:4.000   1st Qu.:3.000   1st Qu.:1984
##  Median :-0.2339   Median :2   Median :5.000   Median :3.000   Median :1989
##  Mean   : 0.0148   Mean   :2   Mean   :4.848   Mean   :2.914   Mean   :1992
##  3rd Qu.: 0.3266   3rd Qu.:2   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:1995
##  Max.   :12.9851   Max.   :9   Max.   :9.000   Max.   :3.000   Max.   :9999
##  NA's   :2235                  NA's   :1       NA's   :1
##       WB4             WDOBLC          MN2             MN5             MN20
##  Min.   :15.00   Min.   :1392   Min.   :1.000   Min.   : 0.00   Min.   :11.00
##  1st Qu.:23.00   1st Qu.:1403   1st Qu.:1.000   1st Qu.: 1.00   1st Qu.:11.00
##  Median :28.00   Median :1409   Median :1.000   Median : 3.00   Median :22.00
##  Mean   :28.51   Mean   :1409   Mean   :1.228   Mean   : 2.94   Mean   :21.21
##  3rd Qu.:33.00   3rd Qu.:1415   3rd Qu.:1.000   3rd Qu.: 4.00   3rd Qu.:23.00
##  Max.   :49.00   Max.   :1422   Max.   :2.000   Max.   :99.00   Max.   :96.00
##
##       PN5             PN17         MN19A              MN19C
##  Min.   :1.000   Min.   :1.00   Length:8467        Length:8467
##  1st Qu.:1.000   1st Qu.:2.00   Class :character   Class :character
##  Median :1.000   Median :2.00   Mode  :character   Mode  :character
##  Mean   :1.496   Mean   :1.93
##  3rd Qu.:2.000   3rd Qu.:2.00
##  Max.   :9.000   Max.   :9.00
##  NA's   :2633    NA's   :2633
##     MN19D              MN19F              MN19G              MN19H
##  Length:8467        Length:8467        Length:8467        Length:8467
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     MN19X              MN19Y                PN9             PN20
##  Length:8467        Length:8467        Min.   :1.000   Min.   :1.000
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.:2.000
##  Mode  :character   Mode  :character   Median :2.000   Median :2.000
##                                        Mean   :1.722   Mean   :1.922
##                                        3rd Qu.:2.000   3rd Qu.:2.000
##                                        Max.   :9.000   Max.   :9.000
##                                        NA's   :6708    NA's   :7593
##      PN25A           PN25B           PN25C           CM17       HHL
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1   Length:8467
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1   Class :character
##  Median :2.000   Median :2.000   Median :2.000   Median :1   Mode  :character
##  Mean   :1.707   Mean   :1.838   Mean   :1.789   Mean   :1
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:1
##  Max.   :2.000   Max.   :2.000   Max.   :2.000   Max.   :1
##
##       HC1A            HC1B            HC2             HH7
##  Min.   : 1.000   Min.   :1.000   Min.   :1.000   Min.   : 1.00
##  1st Qu.: 1.000   1st Qu.:4.000   1st Qu.:1.000   1st Qu.: 8.00
##  Median : 2.000   Median :6.000   Median :1.000   Median :15.00
##  Mean   : 9.887   Mean   :4.925   Mean   :1.259   Mean   :14.24
```

```
##  3rd Qu.: 6.000   3rd Qu.:6.000   3rd Qu.:1.000   3rd Qu.:21.00
##  Max.   :99.000   Max.   :9.000   Max.   :6.000   Max.   :26.00
##  NA's   :1        NA's   :1       NA's   :1
##     helevel           year           ac4           probithpl
##  Min.   :0.000   Min.   :2016   Min.   :0.0000   Min.   :0.000
##  1st Qu.:1.000   1st Qu.:2016   1st Qu.:0.0000   1st Qu.:0.000
##  Median :2.000   Median :2017   Median :0.0000   Median :1.000
##  Mean   :1.957   Mean   :2017   Mean   :0.3743   Mean   :0.689
##  3rd Qu.:3.000   3rd Qu.:2017   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :9.000   Max.   :2018   Max.   :1.0000   Max.   :1.000
##
```

```
# drop irrational year
```

ONE BIG PROBLEM

Only women having live births is last two years are asked about their maternal health seeking behavior.So a selection bias exists.

‘‘‘

Do clustering analysis

```
library(ggplot2)
library(cluster)
library(clustMixType)
library(dplyr)
library(readr)
library(klaR)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(factoextra)
```

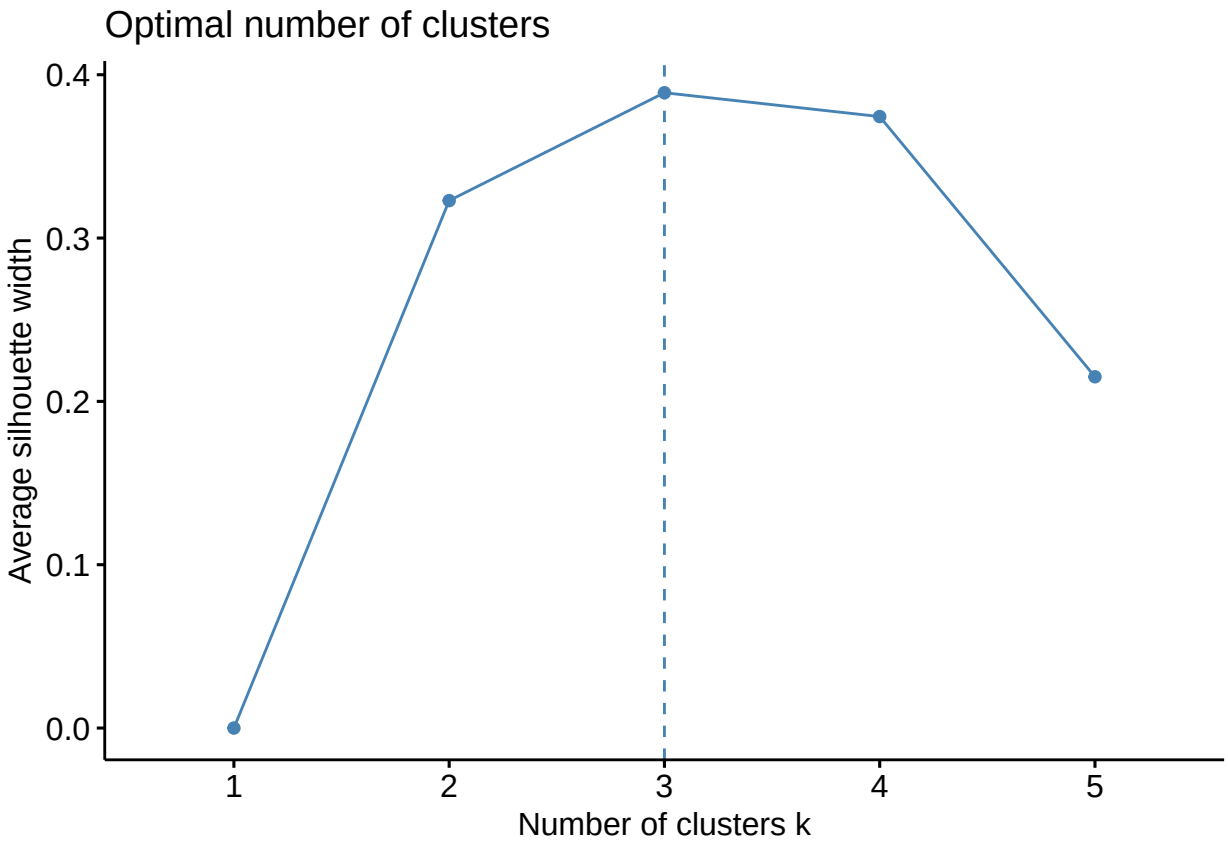
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
wm_clean<-read.csv("D:\\tang_data\\cleandata0503.csv")
hsb<-dplyr::select(wm_clean,MN2,ac4,probithpl,PN25A,
                PN25B,PN25C)
```

Silhouette Index help to determine the optimal cluster numbers

we use five variables to typologzie women’s maternal health seeking behavior: 1)ANC attendence
2)At least 4 times ANC
3)delivery place
4)PNC: examine the cord
5)PNC: take temperature
6)PNC: consel on breastfeeding

```
s<-fviz_nbclust(x=hsb,FUNcluster =  kmodes, method = "silhouette",k.max = 5)
s
```




```
hsb[] <- lapply(hsb, unclass)
```

Therefore we choose 3 clusters

```
set.seed(2)
r<-kmodes(hsb,modes=3)
results<-r$modes
results$MN2[results$MN2==2]<-0
results$PN25A[results$PN25A==2]<-0
results$PN25B[results$PN25B==2]<-0
results$PN25C[results$PN25C==2]<-0
results
```

```
##   MN2 ac4 probithpl PN25A PN25B PN25C
## 1   1   1         1     1     1     1
## 2   0   0         0     0     0     0
## 3   1   0         1     0     0     0
```

MN2: ANC attendence 0: Never Visited 1: Have visited AC4: At least 4 times ANC visits: 0: below 4 times(including never); 1: 4 times and above
probithpl: Delivery place 0: Not in a health professional places 1: Health professional places
PN25A: Examine the cord during the first two days after birth 0:No 1: Yes
PN25B: Examine the temperature during the first two days after birth 0:No 1: Yes

PN25C: Counsel you on breastfeeding during the first two days after birth 0:No 1: Yes

clustering results

Cluster 1: They have fully utilize the maternal health services
Cluster 2: They don't utilize any maternal health services
Cluster 3: They have utilize most of the maternal health services but below the recommended standard

Then I try to use multinomial logit regression to explore demographic, socio-economic status factors influencing their maternal health seeking behavior