

SEN_2015_RIMA

Background and Introduction

Senegal 2015 Resilience Index Measurement and Analysis (RIMA) captures the food security and resilience capacity of households in Senegal in 2015. The data collection in Matam, Senegal aims to obtain an accurate and statistically valid analysis of resilience and food security through the application of the FAO-RIMA (Resilience Index Measurement and Analysis) method. The survey is implemented by the FAO-RAP (Resilience Analysis and Policy) team; the FAO Senegal office and the National Statistics and Demography Agency (ANSD).

The zone selected for the survey is composed by the text following the region of Matam. This area is chronically suffering from droughts and other climatic threats like floods and low rainfall.

Data quality check and Anonymization procedure

First we read in the dataset and perform basic data quality checks as described in the **SDC protocol**.

In most cases, the following data quality checks was conducted:

1. All personal identifying information (i.e. direct identifiers) have been removed.
2. The dataset does not contain any sensitive information which if disclosed would cause harm to the data subjects.
3. The variables names are labeled correctly.
4. Missing (or suppressed) values are coded as “NA”.
5. Reasonable values were confirmed for all numeric variables during cleaning.
6. There are no duplicates.
7. No dataset contain columns of all missing variables.

Read in the dataset

First, we read in the datasets. The datasets includes three parts: consumption, household characteristics and income.

```
## load consumption data
dd <- "/Users/shang/Food and Agriculture Organization/Anidi, Oluwakayode (OCS) - FAM/ESA_RIMA/Senegal/S
data.files <- list.files(dd)

for(i in 1:length(data.files)){
  assign(data.files[i],
    read_dta(file = paste0(dd, "/", data.files[i]))
  )
}
```

```

# load hhchar data
dd <- "/Users/shang/Food and Agriculture Organization/Anidi, Oluwakayode (OCS) - FAM/ESA_RIMA/Senegal/SI

data.files <- list.files(dd)

for(i in 1:length(data.files)){
  assign(data.files[i],
        read_dta(file = paste0(dd, "/", data.files[i]))
  )
}

# load income data
dd <- "/Users/shang/Food and Agriculture Organization/Anidi, Oluwakayode (OCS) - FAM/ESA_RIMA/Senegal/SI

data.files <- list.files(dd)

for(i in 1:length(data.files)){
  assign(data.files[i],
        read_dta(file = paste0(dd, "/", data.files[i]))
  )
}

# find the common variable in all dataset
dfs <- Filter(function(x) is(x, "data.frame"), mget(ls()))

# identify the common variable names
var_names_set <- lapply(dfs, names)
reduce(var_names_set, intersect)

```

```
## [1] "idmen"
```

Then we join all the datasets together by household ID *idmen* (household ID)

```

# join the dataset together
df_all <- reduce(dfs, left_join)

```

```

## Joining, by = "idmen"
## Joining, by = "idmen"
## Joining, by = "idmen"
## Joining, by = "idmen"
## Joining, by = "idmen"
## Joining, by = "idmen"

```

```
## Joining, by = c("idmen", "other")
```

```

## Joining, by = "idmen"
## Joining, by = "idmen"

```

```
## Joining, by = c("idmen", "input_value", "milieu", "crop_div", "crop_share", "rice", "wheat", "millet")
```

```
## Joining, by = "idmen"
```

```
## Joining, by = c("idmen", "sale_liv", "liv_sale", "liv_sale_dol", "liv_sale_dol2")

## Joining, by = c("idmen", "otherincome", "otherincome_dol", "otherincome2", "otherincome_dol2", "inhe

## Joining, by = "idmen"

## Joining, by = c("idmen", "livprod_sale", "livprod_saleimp", "livprod_sale_dol", "liv_train")

## Joining, by = "idmen"

## Joining, by = c("idmen", "rctransfer", "rctrans_need", "rctransfer_val", "rctransfer_dol", "rktransf
```

Then the empty columns and rows are removed.

```
df_all <- df_all %>%
  remove_empty(which = c("rows", "cols"))
```

```
d <- d_org <- df_all
```

Measure the risk in the dataset and apply disclosure limitation methods

We consider two types of disclosure scenarios: 1) matching to other publicly available datasets and 2) spontaneous recognition.

We assume that population registers are already publicly available with the demographic variables gender, age, place of residence (region, urban/rural), religion and other variables such as marital status and variables relating to education and professional status that are also present in our dataset.

The “villages” variable is deemed too detailed and will be removed. The number of female and male household members will be removed because the household size can be calculated by the sum of these two. The ratio of female and male household members will be generated. The variables from “pastoral” to “fisher” measures the same thing as the variable “livelihood”. The variables from “bambara” to “wolof” is the same as “ethnic_group”. “livelihood” and “ethnic_group” are sensitive variables. Thus “pastoral” to “fisher” and “bambara” to “wolof” will not be released. But “livelihood” and “ethnic_group” will be released after anonymization.

```
d <- d %>%
  mutate(fm_ratio = fsize/msize) %>%
  select(-villages, -c(pastoral:fisher), -c(bambara:wolof), -urban)
```

```
keyvars <- c("departement",
  "commune",
  "children",
  "marital_status",
  "agehead",
  "educhead",
  "femhead",
  "iure",
  "hhsz",
  "milieu",
  "livelihood",
  "ethnic_group")
cbind(keyvars, var_label(d[,keyvars]))
```

```
##                keyvars
## departement    "departement"  "departement"
## commune        "commune"      "commune"
## children       "children"      " max number pf children per hh"
## marital_status "marital_status" "marital status of hhead"
## agehead        "agehead"       "age head of hh"
## educhead       "educhead"      "years of education hh head"
## femhead        "femhead"       "female headed hh"
## iure           "iure"          "female hh head de iure - divorce, separe"
## hhsize         "hhsize"        "number of people in hh"
## milieu         "milieu"        "dummy one for rural and zero for urban"
## livelihood     "livelihood"    "livelihood"
## ethnic_group   "ethnic_group"  "ethnic group"
```

The variables that may cause spontaneous recognition are in pramvars listed below

```
pramvars <- c("électricité",
              "irrigation",
              "toilettes",
              "télévision",
              "tracteur",
              "véhicule",
              "moto", "vélo", "bateau")

d[pramvars] <- lapply(d[pramvars], as.factor)

cbind(pramvars, var_label(d[,pramvars]))
```

```
##                pramvars
## électricité    "électricité" NULL
## irrigation     "irrigation"   NULL
## toilettes      "toilettes"    NULL
## télévision     "télévision"   NULL
## tracteur       "tracteur"     NULL
## véhicule       "véhicule"     NULL
## moto           "moto"         NULL
## vélo           "vélo"         NULL
## bateau        "bateau"       NULL
```

We then measure the initial disclosure risk. The number of observations violating 2-anonymity: 403 (97.343%).

```
#d[keyvars] <- lapply(d[keyvars], as.factor)
#d[pramvars] <- lapply(d[pramvars], as.factor)

#sdcInitial<- createSdcObj(dat=d, keyVars=keyvars)

sdcInitial<- createSdcObj(dat=d, keyVars=keyvars, pramVars = pramvars)
sdcInitial
```

```
## The input dataset consists of 414 rows and 301 variables.
## --> Categorical key variables: departement, commune, children, marital_status, agehead, educhead,
## -----
```

```
## Information on categorical key variables:
##
## Reported is the number, mean size and size of the smallest category >0 for recoded variables.
## In parenthesis, the same statistics are shown for the unmodified data.
## Note: NA (missings) are counted as separate categories!
```

##	Key Variable	Number of categories	Mean size
##	departement	3 (3)	138.000 (138.000)
##	commune	16 (16)	25.875 (25.875)
##	children	18 (18)	22.176 (22.176)
##	marital_status	6 (6)	82.000 (82.000)
##	agehead	67 (67)	6.212 (6.212)
##	educhead	8 (8)	44.571 (44.571)
##	femhead	3 (3)	205.000 (205.000)
##	iure	2 (2)	207.000 (207.000)
##	hhsz	31 (31)	13.355 (13.355)
##	milieu	2 (2)	207.000 (207.000)
##	livelihood	6 (6)	27.000 (27.000)
##	ethnic_group	5 (5)	33.750 (33.750)
##	Size of smallest (>0)		
##		84 (84)	
##		11 (11)	
##		1 (1)	
##		3 (3)	
##		1 (1)	
##		1 (1)	
##		88 (88)	
##		55 (55)	
##		1 (1)	
##		189 (189)	
##		2 (2)	
##		1 (1)	

```
## -----
```

```
## Infos on 2/3-Anonymity:
##
## Number of observations violating
## - 2-anonymity: 403 (97.343%)
## - 3-anonymity: 413 (99.758%)
## - 5-anonymity: 414 (100.000%)
##
## -----
```

Recode

We recode the age of household head

```
sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead >= 0 &
sdcInitial@manipKeyVars$agehead <= 18] <- 18
```

```

sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead > 18 &
sdcInitial@manipKeyVars$agehead <= 20] <- 20

sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead > 20 &
sdcInitial@manipKeyVars$agehead <= 30] <- 30

sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead > 30 &
sdcInitial@manipKeyVars$agehead <= 40] <- 40

sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead > 40 &
sdcInitial@manipKeyVars$agehead <= 50] <- 50

sdcInitial@manipKeyVars$agehead[sdcInitial@manipKeyVars$agehead > 50 &
sdcInitial@manipKeyVars$agehead <65] <- 60

sdcInitial <- topBotCoding(obj = sdcInitial, value = 65,
replacement = 65, kind = 'top', column = 'agehead')

```

Recode the education level of household head

```

sdcInitial@manipKeyVars$educhead[sdcInitial@manipKeyVars$educhead >= 0 &
sdcInitial@manipKeyVars$educhead <= 6] <- 6

sdcInitial@manipKeyVars$educhead[sdcInitial@manipKeyVars$educhead > 6 &
sdcInitial@manipKeyVars$educhead <= 9] <- 9

sdcInitial@manipKeyVars$educhead[sdcInitial@manipKeyVars$educhead > 9] <- 10

```

Recode is applied to household size.

```

sdcInitial@manipKeyVars$hhsz[sdcInitial@manipKeyVars$hhsz >= 0 &
sdcInitial@manipKeyVars$hhsz <= 5] <- 5

sdcInitial@manipKeyVars$hhsz[sdcInitial@manipKeyVars$hhsz >=6 &
sdcInitial@manipKeyVars$hhsz <= 10] <- 10

sdcInitial@manipKeyVars$hhsz[sdcInitial@manipKeyVars$hhsz >=11 &
sdcInitial@manipKeyVars$hhsz <= 15] <- 15

sdcInitial@manipKeyVars$hhsz[sdcInitial@manipKeyVars$hhsz >=16] <- 16

```

Topcode number of children

```

sdcInitial <- topBotCoding(sdcInitial, value = 10, replacement = 10, kind = 'top', column = 'children')

```

#Pram

```

set.seed(seed)
sdcInitial <- pram(sdcInitial, strata_variables = "departement", pd = 0.8)
sdcInitial

```

```

## The input dataset consists of 414 rows and 301 variables.
## --> Categorical key variables: departement, commune, children, marital_status, agehead, educhead, :
## -----

## Information on categorical key variables:
##
## Reported is the number, mean size and size of the smallest category >0 for recoded variables.
## In parenthesis, the same statistics are shown for the unmodified data.
## Note: NA (missings) are counted as separate categories!

## Key Variable Number of categories Mean size
## departement 3 (3) 138.000 (138.000)
## commune 16 (16) 25.875 (25.875)
## children 11 (18) 37.700 (22.176)
## marital_status 6 (6) 82.000 (82.000)
## agehead 7 (67) 68.333 (6.212)
## educhead 3 (8) 156.000 (44.571)
## femhead 3 (3) 205.000 (205.000)
## iure 2 (2) 207.000 (207.000)
## hhsz 4 (31) 103.500 (13.355)
## milieu 2 (2) 207.000 (207.000)
## livelihood 6 (6) 27.000 (27.000)
## ethnic_group 5 (5) 33.750 (33.750)
## Size of smallest (>0)
## 84 (84)
## 11 (11)
## 17 (1)
## 3 (3)
## 2 (1)
## 39 (1)
## 88 (88)
## 55 (55)
## 60 (1)
## 189 (189)
## 2 (2)
## 1 (1)

## -----

## Infos on 2/3-Anonymity:
##
## Number of observations violating
## - 2-anonymity: 306 (73.913%) | in original data: 403 (97.343%)
## - 3-anonymity: 380 (91.787%) | in original data: 413 (99.758%)
## - 5-anonymity: 406 (98.068%) | in original data: 414 (100.000%)
##
## -----

## Post-Randomization (PRAM):

## Variable:électricité

## --> final Transition-Matrix:

```

```

##           0           1
## 0 0.9305720 0.06942804
## 1 0.0795005 0.92049950

## Variable:irrigation
##
## --> final Transition-Matrix:

##           0           1
## 0 0.9636696 0.03633036
## 1 0.1443395 0.85566048

## Variable:toilettes
##
## --> final Transition-Matrix:

##           0           1
## 0 0.97380100 0.0261990
## 1 0.01032082 0.9896792

## Variable:télévision
##
## --> final Transition-Matrix:

##           0           1
## 0 0.9276619 0.07233815
## 1 0.1653443 0.83465567

## Variable:tracteur
##
## --> final Transition-Matrix:

##           0
## 0 0.916653

## Variable:véhicule
##
## --> final Transition-Matrix:

##           0           1
## 0 0.9923640 0.007635987
## 1 0.4439781 0.556021873

## Variable:moto
##
## --> final Transition-Matrix:

##           0           1
## 0 0.9816524 0.01834759
## 1 0.3614476 0.63855242

```



```
## Variable:vélo
##
## --> final Transition-Matrix:
```

```
##           0           1
## 0 0.9815078 0.01849216
## 1 0.4318463 0.56815371
```

```
## Variable:bateau
##
## --> final Transition-Matrix:
```

```
##           0           1
## 0 0.9989754 0.00102459
## 1 0.4231556 0.57684435
```

```
##
## Changed observations:
```

```
##      variable nrChanges percChanges
## 1 électricité      29         7.00
## 2 irrigation      19         4.59
## 3 toilettes       4         0.97
## 4 télévision     45        10.87
## 5 tracteur        0         0.00
## 6 véhicule       10         2.42
## 7 moto           18         4.35
## 8 vélo           12         2.90
## 9 bateau         1         0.24
```

```
## -----
```

Local suppression

Finally, local suppression is applied to ensure 2-k anonymity.

```
sdcInitial <- localSuppression(sdcInitial, k = 2)
sdcInitial
```

```
## The input dataset consists of 414 rows and 301 variables.
## --> Categorical key variables: departement, commune, children, marital_status, agehead, educhead, :
## -----
```

```
## Information on categorical key variables:
##
```

```
## Reported is the number, mean size and size of the smallest category >0 for recoded variables.
## In parenthesis, the same statistics are shown for the unmodified data.
## Note: NA (missings) are counted as separate categories!
```

```

##      Key Variable Number of categories      Mean size
##      departement          4 (3)    137.333 (138.000)
##      commune              17 (16)    18.875 (25.875)
##      children             11 (18)    26.000 (22.176)
##      marital_status       6 (6)     77.800 (82.000)
##      agehead              7 (67)    59.667 (6.212)
##      educhead             3 (8)    156.000 (44.571)
##      femhead              3 (3)    202.000 (205.000)
##      iure                 2 (2)    207.000 (207.000)
##      hhsize               5 (31)    96.750 (13.355)
##      milieu               3 (2)    206.500 (207.000)
##      livelihood           6 (6)    26.400 (27.000)
##      ethnic_group         5 (5)    33.750 (33.750)
##      Size of smallest (>0)
##      83 (84)
##      3 (11)
##      10 (1)
##      1 (3)
##      2 (1)
##      39 (1)
##      82 (88)
##      55 (55)
##      52 (1)
##      189 (189)
##      2 (2)
##      1 (1)

## -----

## Infos on 2/3-Anonymity:
##
## Number of observations violating
## - 2-anonymity: 0 (0.000%) | in original data: 403 (97.343%)
## - 3-anonymity: 142 (34.300%) | in original data: 413 (99.758%)
## - 5-anonymity: 338 (81.643%) | in original data: 414 (100.000%)
##
## -----

## Post-Randomization (PRAM):

## Variable:électricité

## --> final Transition-Matrix:

##      0      1
## 0 0.9305720 0.06942804
## 1 0.0795005 0.92049950

## Variable:irrigation
##
## --> final Transition-Matrix:

```

```

##          0          1
## 0 0.9636696 0.03633036
## 1 0.1443395 0.85566048

## Variable:toilettes
##
## --> final Transition-Matrix:

##          0          1
## 0 0.97380100 0.0261990
## 1 0.01032082 0.9896792

## Variable:télévision
##
## --> final Transition-Matrix:

##          0          1
## 0 0.9276619 0.07233815
## 1 0.1653443 0.83465567

## Variable:tracteur
##
## --> final Transition-Matrix:

##          0
## 0 0.916653

## Variable:véhicule
##
## --> final Transition-Matrix:

##          0          1
## 0 0.9923640 0.007635987
## 1 0.4439781 0.556021873

## Variable:moto
##
## --> final Transition-Matrix:

##          0          1
## 0 0.9816524 0.01834759
## 1 0.3614476 0.63855242

## Variable:vélo
##
## --> final Transition-Matrix:

##          0          1
## 0 0.9815078 0.01849216
## 1 0.4318463 0.56815371

```

```

## Variable:bateau
##
## --> final Transition-Matrix:

##           0           1
## 0 0.9989754 0.00102459
## 1 0.4231556 0.57684435

##
## Changed observations:

##      variable nrChanges percChanges
## 1 électricité      29         7.00
## 2 irrigation       19         4.59
## 3 toilettes        4          0.97
## 4 télévision      45        10.87
## 5 tracteur         0          0.00
## 6 véhicule        10         2.42
## 7 moto            18         4.35
## 8 vélo            12         2.90
## 9 bateau          1          0.24

## -----

## Local suppression:

##      KeyVar | Suppressions (#) | Suppressions (%)
##      departement |      2 |      0.483
##      commune |     112 |     27.053
##      children |     117 |     28.261
##      marital_status |     21 |      5.072
##      agehead |     52 |     12.560
##      educhead |      0 |      0.000
##      femhead |      6 |      1.449
##      iure |      0 |      0.000
##      hhsize |     27 |      6.522
##      milieu |      1 |      0.242
##      livelihood |      3 |      0.725
##      ethnic_group |      0 |      0.000

## -----

```

Save anonymized data

The variable labels are added back. Value labels are also added if the variable is altered in the anonymization process. The anonymized data is saved in several format for different software.

```

## Warning in if (cc != class(v_p)) {: the condition has length > 1 and only the
## first element will be used

```

```
## Warning in if (cc != class(v_p)) {: the condition has length > 1 and only the
## first element will be used

## Warning in if (cc != class(v_p)) {: the condition has length > 1 and only the
## first element will be used

## Warning in if (cc != class(v_p)) {: the condition has length > 1 and only the
## first element will be used
```