

Day 3 - Joint Application of TRS and CKM in R

ESTP Course on SDC Methods and Tools for Census 2021

Johannes Gussenbauer and Tobias Enderle

January 29, 2021

0. Preparation

Load R-packages

```
library(data.table)
library(recordSwapping)
library(cellKey)
library(ptable)
```

Load (micro) data

```
dat <- fread("test_data_10k.csv.gz")
```

Be sure, that there are record keys:

```
# first we need a PID (personal identifier; similar to HID)
dat$PID <- 1:nrow(dat)

# draw the record-key
dat$rkey <- ck_generate_rkeys(dat = dat, seed = 123)

# extract PID and the record-key variable
dat_rkey <- dat[, .(PID, rkey)]

# the remaining variables without the record-key
dat <- dat[, !"rkey"]
```

In a future upgrade you won't have to remove the record-key before applying TRS and attach it afterwards.

1. Targeted record swapping (TRS)

1.1 Prepare the data

```
# transform data
# extract X and Y coordinates
dat[,Y_coord:=as.integer(substr(L001000,5,8))]
dat[,X_coord:=as.integer(substr(L001000,10,13))]
dat[1,.(L001000,X_coord,Y_coord)]
```

```
##           L001000 X_coord Y_coord
## 1: 1kmN2793E4656   4656   2793
```

```
dat[,L001000:=NULL] # drop column
dat[,AGE.M:=as.integer(factor(AGE.M))]
dat[,Size:=pmin(5,Size)]
dat[!duplicated(HID),.N,by=.(Size)][order(Size)]
```

```
##      Size      N
## 1:     1 1693
## 2:     2 2719
## 3:     3 1956
## 4:     4 2035
## 5:     5 1597
```

1.2 Set the parameters

```
# parameter
hierarchy <- c("NUTS1","NUTS2")

# hid column
hid <- "HID"

# risk variables
risk_variables <- c("COC.M","POB.M")
k_anonymity <- 3
swaprte <- 0.05
similar <- "Size"
seed <- 2021681
```

1.3 Apply the targeted record swapping

```
dat_swapped <- recordSwap(data = dat, hid = hid,
                           hierarchy = hierarchy,
                           similar = similar,
                           risk_variables = risk_variables,
                           k_anonymity = k_anonymity,
                           swaprte = swaprte,
                           return_swapped_id = TRUE,
                           seed = seed)

dat_swapped
```

```

##      NUTS1 NUTS2 NUTS3  LAU2  HID Size SEX AGE.H AGE.M TPH.H HST EDU SIE
##  1:      1    12   121 121025    1   4   2    1    1   212  11  11   6
##  2:      1    12   121 121025    1   4   2   21    5   212  11   4   6
##  3:      1    12   121 121025    1   4   1   47   10   212  11   5   6
##  4:      1    12   121 121025    1   4   2   55   13   212  11   4   1
##  5:      3    31   312 312074    2   1   1   49   10    11 121   4   1
##  ---
## 30574:    2    22   334 334009 10000    5   1   68   15    4  21   5   6
## 30575:    2    22   334 334009 10000    5   2   71   16    4  21   5   6
## 30576:    2    22   334 334009 10000    5   2   98   21    4  21   5   6
## 30577:    2    22   334 334009 10000    5   2   82   18    4  21  10   6
## 30578:    2    22   334 334009 10000    5   2   76   17    4  21   4   6
##      INCOME_Q COC.ISO3N COC.L COC.M COC.H POB.ISO3N POB.L POB.M POB.H  PID
##  1:          0      276   21   21  2102      40    1    1    1    1
##  2:          1      191   21   21  2105      40    1    1    1    2
##  3:          3      276   21   21  2102      40    1    1    1    3
##  4:          0       40    1    1    1      40    1    1    1    4
##  5:          4       40    1    1    1      40    1    1    1    5
##  ---
## 30574:    2      40    1    1    1      729   22   222 22218 30574
## 30575:    4     348   21   21  2106      40    1    1    1 30575
## 30576:    3     276   21   21  2102      40    1    1    1 30576
## 30577:    3     642   21   21  2101      40    1    1    1 30577
## 30578:    4      40    1    1    1      40    1    1    1 30578
##      Y_coord X_coord HID_swapped
##  1:    2793   4656      1
##  2:    2793   4656      1
##  3:    2793   4656      1
##  4:    2793   4656      1
##  5:    2781   4608      2
##  ---
## 30574:    2688   4394      6688
## 30575:    2688   4394      6688
## 30576:    2688   4394      6688
## 30577:    2688   4394      6688
## 30578:    2688   4394      6688

```

Now attach (merge) the record-key by PID:

```

dat_swapped <- merge(dat_swapped, dat_rkey)
dat_swapped

```

```

##      PID NUTS1 NUTS2 NUTS3  LAU2  HID Size SEX AGE.H AGE.M TPH.H HST EDU
##  1:      1     1    12   121 121025    1   4   2    1    1   212  11  11
##  2:      2     1    12   121 121025    1   4   2   21    5   212  11   4
##  3:      3     1    12   121 121025    1   4   1   47   10   212  11   5
##  4:      4     1    12   121 121025    1   4   2   55   13   212  11   4
##  5:      5     3    31   312 312074    2   1   1   49   10    11 121   4
##  ---
## 30574: 30574    2    22   334 334009 10000    5   1   68   15    4  21   5
## 30575: 30575    2    22   334 334009 10000    5   2   71   16    4  21   5
## 30576: 30576    2    22   334 334009 10000    5   2   98   21    4  21   5
## 30577: 30577    2    22   334 334009 10000    5   2   82   18    4  21  10

```

```

## 30578: 30578      2      22      334 334009 10000      5      2      76      17      4      21      4
##          SIE INCOME_Q COC.IS03N COC.L COC.M COC.H POB.IS03N POB.L POB.M POB.H
## 1:      6          0          276      21      21      2102          40      1      1      1
## 2:      6          1          191      21      21      2105          40      1      1      1
## 3:      6          3          276      21      21      2102          40      1      1      1
## 4:      1          0          40       1       1       1          40      1      1      1
## 5:      1          4          40       1       1       1          40      1      1      1
## ---
## 30574:      6          2          40       1       1       1          729      22      222 22218
## 30575:      6          4          348      21      21      2106          40      1      1      1
## 30576:      6          3          276      21      21      2102          40      1      1      1
## 30577:      6          3          642      21      21      2101          40      1      1      1
## 30578:      6          4          40       1       1       1          40      1      1      1
##          Y_coord X_coord HID_swapped      rkey
## 1:      2793      4656          1 0.2875775
## 2:      2793      4656          1 0.7883051
## 3:      2793      4656          1 0.4089769
## 4:      2793      4656          1 0.8830174
## 5:      2781      4608          2 0.9404673
## ---
## 30574:      2688      4394          6688 0.9132392
## 30575:      2688      4394          6688 0.4338734
## 30576:      2688      4394          6688 0.1884718
## 30577:      2688      4394          6688 0.9772621
## 30578:      2688      4394          6688 0.9600212

```

2. Cell Key method (CKM)

2.1 Design the perturbation table

```
ptabl <- create_cnt_ptable(D = 2, V = 1.08, js = 1, mono = c(T,T,F,T))
```

2.2 Specify the variables (dimensions and hierarchies)

```

d_sex <-
  hier_create(
    nodes = c("1","2"),
    root = "Total"
  );

coc.m_cat <- unique(as.character(dat$COC.M))

d_coc.m <-
  hier_compute(
    inp = coc.m_cat,
    dim_spec = c(1,1,1),
    root = "Total",
    method = "len"
  );

```

2.2 Define the table

```
tab <- ck_setup(  
  x = dat_swapped, # Important: Use the already swapped data  
  rkey = "rkey",  
  dims = list(SEX = d_sex, COC.M = d_coc.m)  
)
```

2.3 Prepare and apply the CKM

```
ptab_input <- ck_params_cnts(ptab = ptab1)  
tab$params_cnts_set(val = ptab_input, v = "total")
```

```
## --> setting perturbation parameters for variable "total"
```

```
# run the perturbation  
tab$perturb(v = "total")
```

```
## Count variable "total" was perturbed.
```

```
# the CKM post-perturbed frequency table of the TRS pre-perturbed micro data  
tab$freqtab(v="total")
```

```
##      SEX COC.M vname   uwc    wc  puwc   pwc  
## 1: Total Total total 30578 30578 30578 30578  
## 2: Total    1 total 25761 25761 25761 25761  
## 3: Total    2 total  4817  4817  4818  4818  
## 4: Total   21 total  2698  2698  2696  2696  
## 5: Total   22 total  2119  2119  2120  2120  
## 6: Total  221 total  1382  1382  1383  1383  
## 7: Total  222 total    91    91    90    90  
## 8: Total  223 total    44    44    44    44  
## 9: Total  224 total    51    51    51    51  
## 10: Total 225 total   544   544   544   544  
## 11: Total 226 total     7     7     7     7  
## 12:    1 Total total 15307 15307 15308 15308  
## 13:    1    1 total 12869 12869 12868 12868  
## 14:    1    2 total  2438  2438  2438  2438  
## 15:    1   21 total  1382  1382  1382  1382  
## 16:    1   22 total  1056  1056  1057  1057  
## 17:    1  221 total   688   688   688   688  
## 18:    1  222 total    42    42    40    40  
## 19:    1  223 total    21    21    20    20  
## 20:    1  224 total    26    26    28    28  
## 21:    1  225 total   277   277   276   276  
## 22:    1  226 total     2     2     3     3  
## 23:    2 Total total 15271 15271 15272 15272  
## 24:    2    1 total 12892 12892 12892 12892  
## 25:    2    2 total  2379  2379  2379  2379
```

## 26:	2	21	total	1316	1316	1316	1316
## 27:	2	22	total	1063	1063	1064	1064
## 28:	2	221	total	694	694	693	693
## 29:	2	222	total	49	49	47	47
## 30:	2	223	total	23	23	22	22
## 31:	2	224	total	25	25	25	25
## 32:	2	225	total	267	267	266	266
## 33:	2	226	total	5	5	6	6
##	SEX	COC.M	vname	uwc	wc	puwc	pwc