

Softwares for two-stage sample design

Comparison of performance

```
In [1]: results <- as.data.frame(list(Variable = rep(NA,24),
                                     Package = rep(NA,24),
                                     Unit = rep(NA,24),
                                     Units = rep(NA,24)))
```

```
In [2]: load("pop.RData")
str(pop)

'data.frame': 2258507 obs. of 13 variables:
 $ region      : Factor w/ 3 levels "north","center",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ province    : Factor w/ 6 levels "north_1","north_2",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ municipality: num 1 1 1 1 1 1 1 1 1 1 ...
 $ id_hh       : Factor w/ 963018 levels "H1","H10","H100",...: 1 1 1 2 3 3 3 3 1114
 1114 ...
 $ id_ind      : int 1 2 3 4 5 6 7 8 9 10 ...
 $ stratum     : Factor w/ 24 levels "1000","2000",...: 12 12 12 12 12 12 12 12 12 1
 2 ...
 $ stratum_label: chr "north_1_6" "north_1_6" "north_1_6" "north_1_6" ...
 $ sex         : int 1 2 1 2 1 1 2 2 1 1 ...
 $ cl_age      : Factor w/ 8 levels "(0,14]","(14,24]",...: 3 7 8 5 4 6 6 4 4 1 ...
 $ active      : num 1 1 0 1 1 1 1 1 1 0 ...
 $ income_hh   : num 30488 30488 30488 21756 29871 ...
 $ unemployed  : num 0 0 0 0 0 0 0 0 0 0 ...
 $ inactive    : num 0 0 1 0 0 0 0 0 0 1 ...
```

1. PractTools

```
In [3]: library(PracTools)
```

```
In [4]: # Probabilities of inclusion I stage
pp <- as.numeric(table(pop$municipality))/nrow(pop)
```

Variable: income_hh

```
In [5]: bw <- BW2stagePPS(pop$income_hh,
                          pp,
                          psuID=pop$municipality)
bw
des1 <- clusOpt2(C1=130,
                C2=1,
                delta=bw[6],
                unit.rv=bw[3],
                k=bw[5],
                CV0=0.02,
                tot.cost=NULL,
                cal.sw=2)
des1
sample_size <- des1$m.opt*des1$n.opt
sample_size
results[1,1] <- "income_hh"
```

```

results[1,2] <- "PractTools"
results[1,3] <- "PSU"
results[1,4] <- round(des1$m.opt)
results[2,1] <- "income_hh"
results[2,2] <- "PractTools"
results[2,3] <- "SSU"
results[2,4] <- round(des1$m.opt*des1$n.opt)

```

B2: 0.0407589298607924 **W2:** 0.795386737128669 **unit relvar:** 0.836017657056984 **B2+W2:** 0.836145666989461 **k:** 1.0001531186948 **delta:** 0.0487462071142995

```

C1 = 130
C2 = 1
delta = 0.04874621
unit relvar = 0.8360177
k = 1.000153
cost = 25499.72
m.opt = 141.4
n.opt = 50.4
CV = 0.02

```

unit relvar: 7126.56

Variable: active

In [6]:

```

bw <- BW2stagePPS(pop$active,
                  pp,
                  psuID=pop$municipality)

bw
des2 <- clusOpt2(C1=180,
                 C2=1,
                 delta=bw[6],
                 unit.rv=bw[3],
                 k=bw[5],
                 CV0=0.03,
                 tot.cost=NULL,
                 cal.sw=2)

des2
sample_size <- des2$m.opt*des2$n.opt
sample_size
results[3,1] <- "active"
results[3,2] <- "PractTools"
results[3,3] <- "PSU"
results[3,4] <- round(des2$m.opt)
results[4,1] <- "active"
results[4,2] <- "PractTools"
results[4,3] <- "SSU"
results[4,4] <- round(des2$m.opt*des2$n.opt)

```

B2: 0.0341600294082881 **W2:** 0.486132985345657 **unit relvar:** 0.520186300252101 **B2+W2:** 0.520293014753945 **k:** 1.000205146698 **delta:** 0.0656553681091469

```

C1 = 180
C2 = 1
delta = 0.06565537
unit relvar = 0.5201863
k = 1.000205
cost = 11214.18
m.opt = 48.6

```

```
n.opt = 50.6
CV = 0.03
```

unit relvar: 2459.16

Variable: inactive

In [7]:

```
# Third variable (inactive)
bw <- BW2stagePPS(pop$inactive,
                  pp,
                  psuID=pop$municipality)

bw
des3 <- clusOpt2(C1=5,
                 C2=1,
                 delta=bw[6],
                 unit.rv=bw[3],
                 k=bw[5],
                 CV0=0.03,
                 tot.cost=NULL,
                 cal.sw=2)

des3
sample_size <- des3$m.opt*des3$n.opt
sample_size
results[5,1] <- "inactive"
results[5,2] <- "PractTools"
results[5,3] <- "PSU"
results[5,4] <- round(des3$m.opt)
results[6,1] <- "inactive"
results[6,2] <- "PractTools"
results[6,3] <- "SSU"
results[6,4] <- round(des3$m.opt*des3$n.opt)
```

B2: 0.00747592548050322 **W2:** 3.58642740200959 **unit relvar:** 3.59308625918809 **B2+W2:**
3.59390332749009 **k:** 1.00022740013544 **delta:** 0.00208016877452412

```
C1 = 5
C2 = 1
delta = 0.002080169
unit relvar = 3.593086
k = 1.000227
cost = 4840.099
m.opt = 89.7
n.opt = 49
CV = 0.03
```

unit relvar: 4395.3

Variable: unemployed

In [8]:

```
# Fourth variable (unemployed)
bw <- BW2stagePPS(pop$unemployed,
                  pp,
                  psuID=pop$municipality)

bw
des4 <- clusOpt2(C1=350, # whatever value, m.opt always exceed number of municipalit
                 C2=1,
                 delta=bw[6],
                 unit.rv=bw[3],
                 k=bw[5],
                 CV0=0.05,
                 tot.cost=NULL,
                 cal.sw=2)
```

```

des4
sample_size <- des4$m.opt*des4$n.opt
sample_size
results[7,1] <- "unemployed"
results[7,2] <- "PractTools"
results[7,3] <- "PSU"
results[7,4] <- round(des4$m.opt)
results[8,1] <- "unemployed"
results[8,2] <- "PractTools"
results[8,3] <- "SSU"
results[8,4] <- round(des4$m.opt*des4$n.opt)

```

B2: 0.889005285809302 **W2:** 6.14659063073452 **unit relvar:** 7.03424949497764 **B2+W2:**
7.03559591654382 **k:** 1.00019140941292 **delta:** 0.126358207087882

```

C1 = 350
C2 = 1
delta = 0.1263582
unit relvar = 7.034249
k = 1.000191
cost = 161905.3
m.opt = 405.6
n.opt = 49.2
CV = 0.05

```

unit relvar: 19955.52

2.R2BEAT

In [9]: `library(R2BEAT)`

Caricamento del pacchetto richiesto: devtools

Caricamento del pacchetto richiesto: usethis

Caricamento del pacchetto richiesto: sampling

Caricamento del pacchetto richiesto: glue

Variable: income_hh

In [10]:

```

# FIRST VARIABLE : income_hh
## Precision constraints
cv <- as.data.frame(list(DOM=c("DOM1"),
                        CV1=c(0.02)))

## -----
## Prepare inputs for allocation
samp_frame <- pop
samp_frame$one <- 1
id_PSU <- "municipality"
id_SSU <- "id_ind"
strata_var <- "one"
target_vars <- c("income_hh")
deff_var <- "stratum"
domain_var <- "one"
delta = 1 # households = survey units
minimum <- 50 # minimum number of SSUs to be interviewed in each selected PSU
deff_sugg <- 1.5 # suggestion for the deff value
inp1 <- prepareInputToAllocation1(samp_frame,

```

```

                                id_PSU,
                                id_SSU,
                                strata_var,
                                target_vars,
                                deff_var,
                                domain_var,
                                minimum,
                                delta,
                                deff_sugg)

inp1$rho
## -----
## Allocation
alloc1 <- beat.2st(stratif = inp1$strata,
                  errors = cv,
                  des_file = inp1$des_file,
                  psu_file = inp1$psu_file,
                  rho = inp1$rho,
                  deff_start = NULL,
                  effst = inp1$effst,
                  minPSUstrat = 2,
                  minnumstrat = 2)
PSUs <- select_PSU(alloc1,plot=F)
results[9,1] <- "income_hh"
results[9,2] <- "R2BEAT"
results[9,3] <- "PSU"
results[9,4] <- PSUs$PSU_stats$PSU[nrow(PSUs$PSU_stats)]
results[10,1] <- "income_hh"
results[10,2] <- "R2BEAT"
results[10,3] <- "SSU"
results[10,4] <- PSUs$PSU_stats$SSU[nrow(PSUs$PSU_stats)]

```

Calculating strata...
 Computations are being done on population data

Number of strata: 1
 ... of which with only one unit: 0
 Calculating rho in strata...
 Stratum 1
 A data.frame: 1 × 3

STRATUM RHO_AR1 RHO_NAR1

	<fct>	<dbl>	<dbl>
1	1	0.04875369	

	iterations	PSU_SR	PSU	NSR	PSU	Total	SSU
1	0	0	0		0	2089	
2	1	8	32		40	5645	
3	2	31	62		93	4652	
4	3	21	58		79	4980	

Variable: active

In [11]:

```

# SECOND VARIABLE : active

## Precision constraints
cv <- as.data.frame(list(DOM=c("DOM1"),
                        CV1=c(0.03)))
## -----
## Prepare inputs for allocation
target_vars <- c("active")
inp2 <- prepareInputToAllocation1(samp_frame,
                                id_PSU,
                                id_SSU,

```

```

                                strata_var,
                                target_vars,
                                deff_var,
                                domain_var,
                                minimum,
                                delta,
                                deff_sugg)

inp2$rho
## -----
## Allocation
alloc2 <- beat.2st(stratif = inp2$strata,
                  errors = cv,
                  des_file = inp2$des_file,
                  psu_file = inp2$psu_file,
                  rho = inp2$rho,
                  deff_start = NULL,
                  effst = inp2$effst,
                  minPSUstrat = 2,
                  minnumstrat = 2)
PSUs <- select_PSU(alloc2,plot=F)
results[11,1] <- "active"
results[11,2] <- "R2BEAT"
results[11,3] <- "PSU"
results[11,4] <- PSUs$PSU_stats$PSU[nrow(PSUs$PSU_stats)]
results[12,1] <- "active"
results[12,2] <- "R2BEAT"
results[12,3] <- "SSU"
results[12,4] <- PSUs$PSU_stats$SSU[nrow(PSUs$PSU_stats)]

```

Calculating strata...
 Computations are being done on population data

Number of strata: 1
 ... of which with only one unit: 0
 Calculating rho in strata...
 Stratum 1

A data.frame: 1 × 3

STRATUM RHO_AR1 RHO_NAR1

	<fct>	<dbl>	<dbl>
1	1	0.06566887	

	iterations	PSU_SR	PSU	NSR	PSU	Total	SSU
1	0	0	0		0	578	
2	1	0	12		12	2436	
3	2	11	34		45	1822	
4	3	6	30		36	1978	
5	4	7	30		37	1936	

Variable: inactive

In [12]:

```

## Precision constraints
cv <- as.data.frame(list(DOM=c("DOM1"),
                        CV1=c(0.03)))
## -----
## Prepare inputs for allocation
target_vars <- c("inactive")
inp3 <- prepareInputToAllocation1(samp_frame,
                                id_PSU,
                                id_SSU,
                                strata_var,
                                target_vars,
                                deff_var,

```

```

domain_var,
minimum,
delta,
deff_sugg)

inp3$rho

## -----
## Allocation
alloc3 <- beat.2st(stratif = inp3$strata,
  errors = cv,
  des_file = inp3$des_file,
  psu_file = inp3$psu_file,
  rho = inp3$rho,
  deff_start = NULL,
  effst = inp3$effst,
  minPSUstrat = 2,
  minnumstrat = 2)

PSUs <- select_PSU(alloc3, plot=F)
results[13,1] <- "inactive"
results[13,2] <- "R2BEAT"
results[13,3] <- "PSU"
results[13,4] <- PSUs$PSU_stats$PSU[nrow(PSUs$PSU_stats)]
results[14,1] <- "inactive"
results[14,2] <- "R2BEAT"
results[14,3] <- "SSU"
results[14,4] <- PSUs$PSU_stats$SSU[nrow(PSUs$PSU_stats)]

```

Calculating strata...
Computations are being done on population data

```
Number of strata: 1
... of which with only one unit: 0
Calculating rho in strata...
Stratum 1
```

A data.frame: 1 × 3

STRATUM RHO_AR1 RHO_NAR1

<fct>	<dbl>	<dbl>
1	1	0.002080643

	iterations	PSU_SR	PSU	NSR	PSU	Total	SSU
1	0	0		0		0	3986
2	1	14		54		68	4244

Variable: unemployed

In [13]:

[illegible]

```
## Allocation
alloc4 <- beat.2st(stratif = inp4$strata,
  errors = cv,
  des_file = inp4$des_file,
  psu_file = inp4$psu_file,
  rho = inp4$rho,
  deft_start = NULL,
  effst = inp4$effst,
  minPSUstrat = 2,
  minnumstrat = 2)
PSUs <- select_PSU(alloc4,plot=F)
results[15,1] <- "unemployed"
results[15,2] <- "R2BEAT"
results[15,3] <- "PSU"
results[15,4] <- PSUs$PSU_stats$PSU[nrow(PSUs$PSU_stats)]
results[16,1] <- "unemployed"
results[16,2] <- "R2BEAT"
results[16,3] <- "SSU"
results[16,4] <- PSUs$PSU_stats$SSU[nrow(PSUs$PSU_stats)]
```

Calculating strata...
 Computations are being done on population data

Number of strata: 1
 ... of which with only one unit: 0
 Calculating rho in strata...
 Stratum 1
 A data.frame: 1 × 3

STRATUM RHO_AR1 RHO_NAR1

	<fct>	<dbl>	<dbl>
1	1	1	0.1263824

	iterations	PSU_SR	PSU	NSR	PSU	Total	SSU
1	0	0	0	0	0	2811	
2	1	11	40	51	14386		
3	2	75	114	189	9074		
4	3	45	88	133	10552		
5	4	51	98	149	10198		

3.samplesize4surveys

In [14]: `library("samplesize4surveys")`

Caricamento del pacchetto richiesto: TeachingSampling

Caricamento del pacchetto richiesto: dplyr

Caricamento pacchetto: 'dplyr'

Il seguente oggetto è mascherato da 'package:glue':

`collapse`

I seguenti oggetti sono mascherati da 'package:stats':

`filter`, `lag`

I seguenti oggetti sono mascherati da 'package:base':

intersect, setdiff, setequal, union

Caricamento del pacchetto richiesto: magrittr

Caricamento del pacchetto richiesto: timeDate

```
In [15]: pop_strata <- as.numeric(table(pop$stratum))
PSU <- length(unique(pop$municipality))
rho_income_hh <- inp1$rho$RHO_NAR1
rho_active <- inp2$rho$RHO_NAR1
rho_inactive<- inp3$rho$RHO_NAR1
rho_unemployed <- inp4$rho$RHO_NAR1
```

Variable: income_hh

```
In [16]: # First variable (income_hh)
a1 <- ss2s4m(N = nrow(pop),
  mu = mean(pop$income_hh),
  sigma = sd(pop$income_hh),
  # conf = 0.95,
  delta = 0.02 * 1.96,
  M = PSU,
  to = 50,
  rho = rho_income_hh)
a1[50,]
results[16,4] <- alloc4$iterations[nrow(alloc4$iterations),5]
results[17,1] <- "income_hh"
results[17,2] <- "samplesize4surveys"
results[17,3] <- "PSU"
results[17,4] <- a1[50,2]
results[18,1] <- "income_hh"
results[18,2] <- "samplesize4surveys"
results[18,3] <- "SSU"
results[18,4] <- a1[50,4]
```

A data.frame: 1 × 4

	Deff	nl	m	n2s
	<dbl>	<dbl>	<int>	<dbl>
50	3.388931	142	50	7061

Variable: active

```
In [17]: # Second variable (active)
a2 <- ss2s4p(N = nrow(pop),
  P = as.numeric(table(pop$active))[2]/nrow(pop),
  # conf = 0.95,
  delta = 0.03 * 1.96,
  M = PSU,
  to = 50,
  rho = rho_active)
a2[50,]
results[19,1] <- "active"
results[19,2] <- "samplesize4surveys"
results[19,3] <- "PSU"
```

```

results[19,4] <- a2[50,2]
results[20,1] <- "active"
results[20,2] <- "samplesize4surveys"
results[20,3] <- "SSU"
results[20,4] <- a2[50,4]

```

A data.frame: 1 × 4

	Deff	nl	m	n2s
	<dbl>	<dbl>	<int>	<dbl>
50	4.217774	49	50	2436

Variable: inactive

In [18]:

```

a3 <- ss2s4p(N = nrow(pop),
  P = as.numeric(table(pop$inactive))[2]/nrow(pop),
  # conf = 0.95,
  delta = 0.03 * 1.96,
  M = PSU,
  to = 50,
  rho = rho_inactive)
a3[50,]
results[21,1] <- "inactive"
results[21,2] <- "samplesize4surveys"
results[21,3] <- "PSU"
results[21,4] <- a3[50,2]
results[22,1] <- "inactive"
results[22,2] <- "samplesize4surveys"
results[22,3] <- "SSU"
results[22,4] <- a3[50,4]

```

A data.frame: 1 × 4

	Deff	nl	m	n2s
	<dbl>	<dbl>	<int>	<dbl>
50	1.101951	88	50	4391

Variable: unemployed

In [19]:

```

# Fourth variable (unemployed)
a4 <- ss2s4p(N = nrow(pop),
  P = as.numeric(table(pop$unemployed))[2]/nrow(pop),
  # conf = 0.95,
  delta = 0.05 * 1.96,
  M = PSU,
  to = 50,
  rho = rho_unemployed)
a4[50,]
results[23,1] <- "unemployed"
results[23,2] <- "samplesize4surveys"
results[23,3] <- "PSU"
results[23,4] <- a4[50,2]
results[24,1] <- "unemployed"
results[24,2] <- "samplesize4surveys"
results[24,3] <- "SSU"
results[24,4] <- a4[50,4]

```

A data.frame: 1 × 4

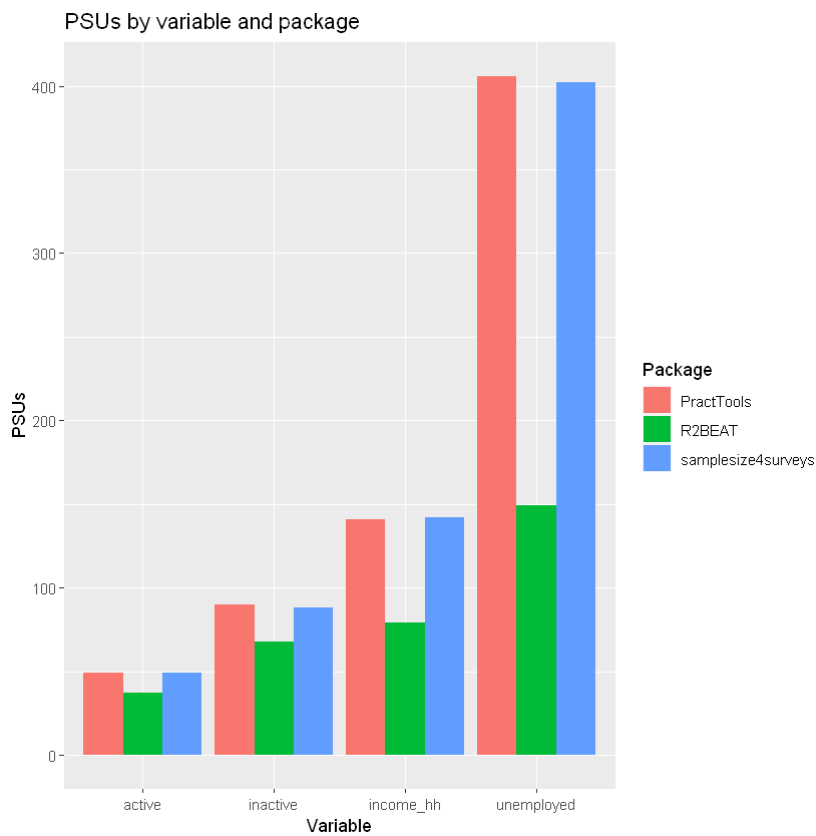
	Deff	nl	m	n2s
	<dbl>	<dbl>	<int>	<dbl>
50	7.19274	402	50	20058

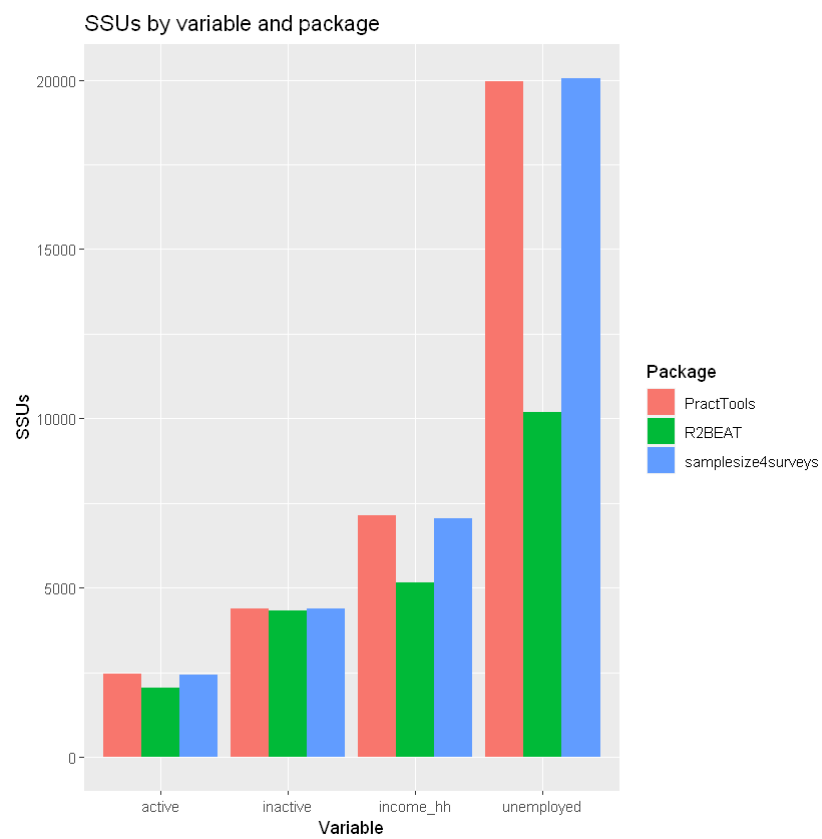
4. Plotting results

In [20]:

```
library(ggplot2)

p <- ggplot(data=results[results$Unit=="PSU",], aes(x=Variable, y=Units, fill=Packag
  geom_bar(stat="identity", position=position_dodge(),)
p + labs(title="PSUs by variable and package",
  x="Variable", y = "PSUs")
p <- ggplot(data=results[results$Unit=="SSU",], aes(x=Variable, y=Units, fill=Packag
  geom_bar(stat="identity", position=position_dodge(),)
p + labs(title="SSUs by variable and package",
  x="Variable", y = "SSUs")
```





In []: