

Workflow example with R2BEAT

Scenario 2

Together with a sampling frame containing the units of the population of reference, also a previous round of the sampling survey to be planned is available

```
In [35]: # Install last version of R2BEAT
#devtools::install_github("barcaroli/R2BEAT", dependencies = FALSE)
library(R2BEAT)
```

```
In [36]: packageVersion("R2BEAT")
```

```
[1] '1.0.4'
```

```
In [37]: ## Sampling frame
load("pop.RData")
```

```
In [38]: ## Sample data
load("sample.RData")
```

Analysis of sampled data

```
In [39]: # Install ReGenesees
#devtools::install_github("DiegoZardetto/ReGenesees")
library(ReGenesees)
```

```
In [40]: ## Sample design description
sample$stratum_2 <- as.factor(sample$stratum_2)
sample.des <- e.svydesign(sample,
  ids= ~ municipality + id_hh,
  strata = ~ stratum_2,
  weights = ~ weight,
  self.rep.str = ~ SR,
  check.data = TRUE)
```

```
# Empty levels found in factors: id_hh
# Empty levels have been dropped!
```

```
In [41]: ## Find and collapse lonely strata
ls <- find.lon.strata(sample.des)
sample.des <- collapse.strata(sample.des)
```

```
# All lonely strata (112) successfully collapsed!
```

```
In [42]: ## Calibration with known totals
totals <- pop.template(sample.des,
  calmodel = ~ sex : cl_age,
  partition = ~ region)
totals <- fill.template(pop, totals, mem.frac = 10)
sample.cal <- e.calibrate(sample.des,
```

```

totals,
calmodel = ~ sex : cl_age,
partition = ~ region,
calfun = "logit",
bounds = c(0.3, 2.6),
aggregate.stage = 2,
force = FALSE)

```

Coherence check between 'universe' and 'template': OK

Preparation of inputs for allocation steps

In [43]:

```

## Preparation of inputs for allocation steps
samp_frame <- pop
RGdes <- sample.des
RGcal <- sample.cal
strata_vars <- c("stratum")
target_vars <- c("income_hh",
               "active",
               "inactive",
               "unemployed")
weight_var <- "weight"
deff_vars <- "stratum"
id_PSU <- c("municipality")
id_SSU <- c("id_hh")
domain_vars <- c("region")
delta <- 1
minimum <- 50

inp <- prepareInputToAllocation2(
  samp_frame, # sampling frame
  RGdes,      # ReGenesees design object
  RGcal,      # ReGenesees calibrated object
  id_PSU,     # identification variable of PSUs
  id_SSU,     # identification variable of SSUs
  strata_vars, # strata variables
  target_vars, # target variables
  deff_vars,  # deff variables
  domain_vars, # domain variables
  delta,      # Average number of SSUs for each selection unit
  minimum     # Minimum number of SSUs to be selected in each PSU
)

```

In [44]:

```
head(inp$strata)
```

A data.frame: 6 × 15

	stratum	STRATUM	N	M1	M2	M3	M4	S1	S2	
	<fct>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	1000	1000	196769	23339.70	0.6801679	0.2127596	0.10707247	16543.72	0.4664113	0.4
2	10000	10000	106057	29340.38	0.7793318	0.2047430	0.01592524	25031.44	0.4146972	0.4
3	11000	11000	205839	27822.70	0.7814228	0.2029522	0.01562493	26050.40	0.4132810	0.4
4	12000	12000	57606	23110.90	0.7632522	0.2079530	0.02879485	15405.51	0.4250862	0.4
5	13000	13000	102801	28185.38	0.7516670	0.2142238	0.03410920	24393.71	0.4320460	0.4
6	14000	14000	84077	24787.12	0.7537232	0.2131530	0.03312385	17403.58	0.4308417	0.4

In [45]:

```
head(inp$deff)
```

A data.frame: 6 × 7

	stratum	STRATUM	DEFF1	DEFF2	DEFF3	DEFF4	b_nar
	<fct>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1000	1000	1.002141	1.003487	1.018508	0.998091	254.50000
2	10000	10000	1.019820	1.029362	1.010320	1.000982	178.83333
3	11000	11000	1.128662	1.036882	1.002039	1.115932	52.07500
4	12000	12000	3.233942	0.978419	1.202842	0.639357	49.42857
5	13000	13000	1.063373	1.056811	1.015756	1.048938	1285.00000
6	14000	14000	1.018801	1.003173	1.002272	1.013573	263.50000

In [46]:

```
head(inp$effst)
```

A data.frame: 6 × 6

	stratum	STRATUM	EFFST1	EFFST2	EFFST3	EFFST4
	<fct>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	1000	1000	0.9875397	0.8647755	0.7565498	1.0033213
2	10000	10000	0.9948599	0.9076545	0.8982699	1.0054137
3	11000	11000	0.9765404	0.8136085	0.7835224	0.9925166
4	12000	12000	1.0145565	0.9113590	0.9126909	1.0007101
5	13000	13000	1.0045911	0.9263170	0.9180502	0.9942647
6	14000	14000	1.0016745	0.9471318	0.9375788	0.9967146

In [47]:

```
head(inp$rho)
```

A data.frame: 6 × 9

	STRATUM	RHO_AR1	RHO_NAR1	RHO_AR2	RHO_NAR2	RHO_AR3	RHO_NAR3	RHO_AR4
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1000	1	0.000008445759	1	0.00001375542	1	0.000073009862	1
2	10000	1	0.000111452671	1	0.00016510965	1	0.000058031865	1
3	11000	1	0.002519079785	1	0.00072211454	1	0.000039921684	1
4	12000	1	0.046128595870	1	-0.00044562537	1	0.004188477876	1
5	13000	1	0.000049355919	1	0.00004424533	1	0.000012271028	1
6	14000	1	0.000071622857	1	0.00001208762	1	0.000008655238	1

In [48]:

```
head(inp$psu_file)
```

A data.frame: 6 × 3

	PSU_ID	STRATUM	PSU_MOS
	<dbl>	<fct>	<dbl>
1	309	1000	50845
2	330	1000	146162
3	292	2000	24794
4	293	2000	19609
5	300	2000	13897
6	304	2000	36195

In [49]:

```
head(inp$des_file)
```

A data.frame: 6 × 4

	STRATUM	STRAT_MOS	DELTA	MINIMUM
	<fct>	<dbl>	<dbl>	<dbl>
1	1000	197007	1	50
2	2000	261456	1	50
3	3000	115813	1	50
4	4000	17241	1	50
5	5000	101067	1	50
6	6000	47218	1	50

Allocation

In [50]:

```
## Precision constraints
cv <- as.data.frame(list(DOM=c("DOM1", "DOM2"),
                        CV1=c(0.02, 0.03),
                        CV2=c(0.03, 0.06),
                        CV3=c(0.03, 0.06),
                        CV4=c(0.03, 0.06)))

cv
```

A data.frame: 2 × 5

DOM	CV1	CV2	CV3	CV4
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
DOM1	0.02	0.03	0.03	0.03
DOM2	0.03	0.06	0.06	0.06

In [51]:

```
alloc <- beat.2st(stratif = inp$strata,
                  errors = cv,
                  des_file = inp$des_file,
                  psu_file = inp$psu_file,
                  rho = inp$rho,
                  deft_start = NULL,
                  effst = inp$effst,
                  epsilon1 = 5,
```

```

mmdiff_deft = 1,
maxi = 15,
epsilon = 10^(-11),
minnumstrat = 2,
maxiter = 200,
maxiter1 = 25)

```

```

iterations PSU_SR PSU NSR PSU Total SSU
1          0      0      0      0 13512
2          1     78     67     145 13209
3          2     44    124     168 13016
4          3     43    123     166 13011

```

Selection of PSUs (I stage)

In [52]:

```

allocat <- alloc$alloc[-nrow(alloc$alloc),]
set.seed(1234)
sample_2st <- StratSel(dataPop= inp$psu_file,
                        idpsu= ~ PSU_ID,
                        dom= ~ STRATUM,
                        final_pop= ~ PSU_MOS,
                        size= ~ PSU_MOS,
                        PSUsamplestratum= 1,
                        min_sample= minimum,
                        min_sample_index= FALSE,
                        dataAll=allocat,
                        domAll= ~ factor(STRATUM),
                        f_sample= ~ ALLOC,
                        planned_min_sample= NULL,
                        launch= F)

```

In [53]:

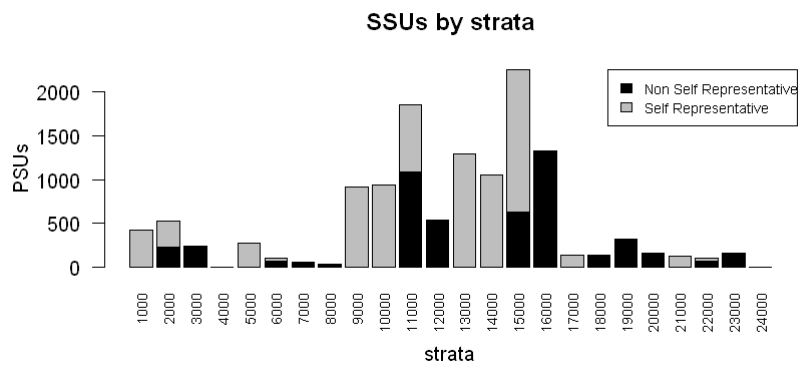
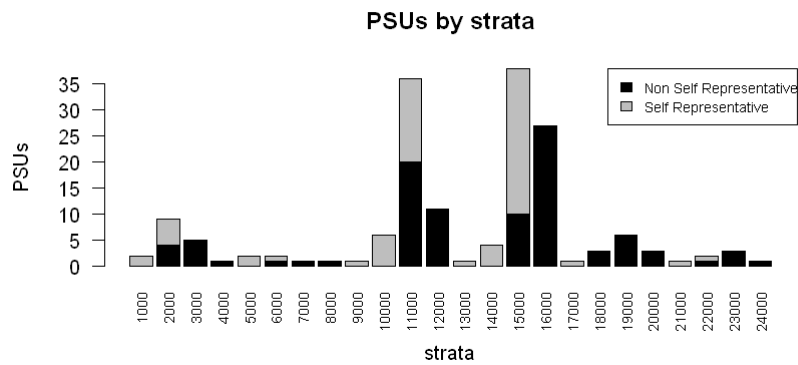
```
sample_2st[[2]]
```

A data.frame: 26 × 6

Domain	SRdom	nSRdom	SRdom+nSRdom	SR_PSU_final_sample_unit	NSR_PSU_final_sample_unit
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1000	2	0	2	423	0
2000	5	4	9	288	235
3000	0	5	5	0	247
4000	0	1	1	0	2
5000	2	0	2	281	0
6000	1	1	2	43	66
7000	0	1	1	0	56
8000	0	1	1	0	35
9000	1	0	1	911	0
10000	6	0	6	936	0
11000	16	20	36	761	1091
12000	0	11	11	0	537
13000	1	0	1	1298	0
14000	4	0	4	1049	0

Domain	SRdom	nSRdom	SRdom+nSRdom	SR_PSU_final_sample_unit	NSR_PSU_final_sample_unit
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
15000	28	10	38	1629	627
16000	0	27	27	0	1333
17000	1	0	1	141	0
18000	0	3	3	0	134
19000	0	6	6	0	320
20000	0	3	3	0	166
21000	1	0	1	130	0
22000	1	1	2	41	68
23000	0	3	3	0	165
24000	0	1	1	0	2
Total	69	98	167	7931	5084
Mean				330	212

```
In [54]: ## Plot of allocation (PSUs and SSUs)
des <- sample_2st[[2]]
des2 <- NULL
des2$strata <- c(des$Domain[1:24],des$Domain[1:24])
des2$SR <- c(rep("SR",24),rep("nSR",24))
des2$PSU <- as.numeric(c(des$SRdom[1:24],des$nSRdom[1:24]))
des2$SSU <- as.numeric(c(des$SR_PSU_final_sample_unit[1:24],des$NSR_PSU_final_sample_unit[1:24]))
des2 <- as.data.frame(des2)
des2$strata <- as.numeric(des2$strata)
par(mfrow=c(2, 1))
barplot(PSU~SR+strata, data=des2,
        main = "PSUs by strata",
        xlab = "strata", ylab = "PSUs",
        col = c("black", "grey"),
        # beside = TRUE,
        las=2,
        cex.names=0.7)
legend("topright",
       legend = c("Non Self Representative","Self Representative"),cex = 0.7,
       fill = c("black", "grey"))
barplot(SSU~SR+strata, data=des2,
        main = "SSUs by strata",
        xlab = "strata", ylab = "PSUs",
        col = c("black", "grey"),
        # beside = TRUE,
        las=2,
        cex.names=0.7)
legend("topright",
       legend = c("Non Self Representative","Self Representative"),cex = 0.7,
       fill = c("black", "grey"))
```



Selection of SSUs (II stage)

In [55]:

```
selected_PSU <- sample_2st[[4]]
selected_PSU <- selected_PSU[selected_PSU$PSU_final_sample_unit > 0,]
samp <- select_SSU(df=pop,
  PSU_code="municipality",
  SSU_code="id_ind",
  PSU_sampled=selected_PSU[selected_PSU$Sampled_PSU==1,],
  verbose=TRUE)
```

```
PSU = 1 *** Selected SSU = 48
PSU = 2 *** Selected SSU = 115
PSU = 3 *** Selected SSU = 58
PSU = 4 *** Selected SSU = 43
PSU = 5 *** Selected SSU = 911
PSU = 6 *** Selected SSU = 52
PSU = 7 *** Selected SSU = 167
PSU = 8 *** Selected SSU = 126
PSU = 9 *** Selected SSU = 63
PSU = 10 *** Selected SSU = 66
PSU = 11 *** Selected SSU = 44
PSU = 12 *** Selected SSU = 56
PSU = 13 *** Selected SSU = 55
PSU = 14 *** Selected SSU = 45
PSU = 15 *** Selected SSU = 42
PSU = 16 *** Selected SSU = 60
PSU = 17 *** Selected SSU = 55
PSU = 18 *** Selected SSU = 42
PSU = 19 *** Selected SSU = 53
PSU = 20 *** Selected SSU = 55
PSU = 21 *** Selected SSU = 42
PSU = 22 *** Selected SSU = 41
PSU = 23 *** Selected SSU = 43
PSU = 24 *** Selected SSU = 138
PSU = 25 *** Selected SSU = 93
PSU = 26 *** Selected SSU = 41
PSU = 27 *** Selected SSU = 49
```

PSU = 28	***	Selected	SSU = 54
PSU = 29	***	Selected	SSU = 297
PSU = 30	***	Selected	SSU = 47
PSU = 31	***	Selected	SSU = 49
PSU = 32	***	Selected	SSU = 47
PSU = 33	***	Selected	SSU = 50
PSU = 34	***	Selected	SSU = 39
PSU = 35	***	Selected	SSU = 62
PSU = 36	***	Selected	SSU = 49
PSU = 37	***	Selected	SSU = 47
PSU = 38	***	Selected	SSU = 56
PSU = 39	***	Selected	SSU = 36
PSU = 40	***	Selected	SSU = 63
PSU = 41	***	Selected	SSU = 58
PSU = 42	***	Selected	SSU = 64
PSU = 43	***	Selected	SSU = 51
PSU = 44	***	Selected	SSU = 57
PSU = 45	***	Selected	SSU = 51
PSU = 46	***	Selected	SSU = 48
PSU = 47	***	Selected	SSU = 44
PSU = 48	***	Selected	SSU = 71
PSU = 49	***	Selected	SSU = 47
PSU = 50	***	Selected	SSU = 49
PSU = 51	***	Selected	SSU = 54
PSU = 52	***	Selected	SSU = 52
PSU = 53	***	Selected	SSU = 46
PSU = 54	***	Selected	SSU = 45
PSU = 55	***	Selected	SSU = 97
PSU = 56	***	Selected	SSU = 51
PSU = 57	***	Selected	SSU = 195
PSU = 58	***	Selected	SSU = 45
PSU = 59	***	Selected	SSU = 55
PSU = 60	***	Selected	SSU = 57
PSU = 61	***	Selected	SSU = 76
PSU = 62	***	Selected	SSU = 43
PSU = 63	***	Selected	SSU = 49
PSU = 64	***	Selected	SSU = 51
PSU = 65	***	Selected	SSU = 51
PSU = 66	***	Selected	SSU = 40
PSU = 67	***	Selected	SSU = 51
PSU = 68	***	Selected	SSU = 51
PSU = 69	***	Selected	SSU = 49
PSU = 70	***	Selected	SSU = 53
PSU = 71	***	Selected	SSU = 51
PSU = 72	***	Selected	SSU = 47
PSU = 73	***	Selected	SSU = 45
PSU = 74	***	Selected	SSU = 44
PSU = 75	***	Selected	SSU = 75
PSU = 76	***	Selected	SSU = 53
PSU = 77	***	Selected	SSU = 47
PSU = 78	***	Selected	SSU = 50
PSU = 79	***	Selected	SSU = 96
PSU = 80	***	Selected	SSU = 85
PSU = 81	***	Selected	SSU = 58
PSU = 82	***	Selected	SSU = 76
PSU = 83	***	Selected	SSU = 106
PSU = 84	***	Selected	SSU = 61
PSU = 85	***	Selected	SSU = 46
PSU = 86	***	Selected	SSU = 41
PSU = 87	***	Selected	SSU = 236
PSU = 88	***	Selected	SSU = 51
PSU = 89	***	Selected	SSU = 70
PSU = 90	***	Selected	SSU = 53
PSU = 91	***	Selected	SSU = 50
PSU = 92	***	Selected	SSU = 188
PSU = 93	***	Selected	SSU = 64
PSU = 94	***	Selected	SSU = 55
PSU = 95	***	Selected	SSU = 430
PSU = 96	***	Selected	SSU = 65

PSU = 97	***	Selected SSU = 49
PSU = 98	***	Selected SSU = 57
PSU = 99	***	Selected SSU = 48
PSU = 100	***	Selected SSU = 52
PSU = 101	***	Selected SSU = 44
PSU = 102	***	Selected SSU = 40
PSU = 103	***	Selected SSU = 45
PSU = 104	***	Selected SSU = 1298
PSU = 105	***	Selected SSU = 43
PSU = 106	***	Selected SSU = 42
PSU = 107	***	Selected SSU = 32
PSU = 108	***	Selected SSU = 60
PSU = 109	***	Selected SSU = 72
PSU = 110	***	Selected SSU = 55
PSU = 111	***	Selected SSU = 48
PSU = 112	***	Selected SSU = 58
PSU = 113	***	Selected SSU = 55
PSU = 114	***	Selected SSU = 45
PSU = 115	***	Selected SSU = 72
PSU = 116	***	Selected SSU = 25
PSU = 117	***	Selected SSU = 48
PSU = 118	***	Selected SSU = 59
PSU = 119	***	Selected SSU = 66
PSU = 120	***	Selected SSU = 49
PSU = 121	***	Selected SSU = 55
PSU = 122	***	Selected SSU = 55
PSU = 123	***	Selected SSU = 38
PSU = 124	***	Selected SSU = 69
PSU = 125	***	Selected SSU = 50
PSU = 126	***	Selected SSU = 39
PSU = 127	***	Selected SSU = 72
PSU = 128	***	Selected SSU = 72
PSU = 129	***	Selected SSU = 51
PSU = 130	***	Selected SSU = 109
PSU = 131	***	Selected SSU = 46
PSU = 132	***	Selected SSU = 51
PSU = 133	***	Selected SSU = 55
PSU = 134	***	Selected SSU = 59
PSU = 135	***	Selected SSU = 52
PSU = 136	***	Selected SSU = 314
PSU = 137	***	Selected SSU = 48
PSU = 138	***	Selected SSU = 2
PSU = 139	***	Selected SSU = 68
PSU = 140	***	Selected SSU = 60
PSU = 141	***	Selected SSU = 47
PSU = 142	***	Selected SSU = 35
PSU = 143	***	Selected SSU = 66
PSU = 144	***	Selected SSU = 43
PSU = 145	***	Selected SSU = 56
PSU = 146	***	Selected SSU = 6
PSU = 147	***	Selected SSU = 275
PSU = 148	***	Selected SSU = 54
PSU = 149	***	Selected SSU = 52
PSU = 150	***	Selected SSU = 56
PSU = 151	***	Selected SSU = 56
PSU = 152	***	Selected SSU = 51
PSU = 153	***	Selected SSU = 48
PSU = 154	***	Selected SSU = 56
PSU = 155	***	Selected SSU = 51
PSU = 156	***	Selected SSU = 141
PSU = 157	***	Selected SSU = 44
PSU = 158	***	Selected SSU = 57
PSU = 159	***	Selected SSU = 39
PSU = 160	***	Selected SSU = 56
PSU = 161	***	Selected SSU = 2
PSU = 162	***	Selected SSU = 54
PSU = 163	***	Selected SSU = 60
PSU = 164	***	Selected SSU = 130
PSU = 165	***	Selected SSU = 41

```

PSU = 166 *** Selected SSU = 68
PSU = 167 *** Selected SSU = 51
-----
Total PSU = 167
Total SSU = 13015
-----

```

In [56]:

```

nrow(samp)
sum(allocat$ALLOC)

```

13015

13011

In [57]:

```

nrow(pop)
sum(samp$weight)

```

2258507

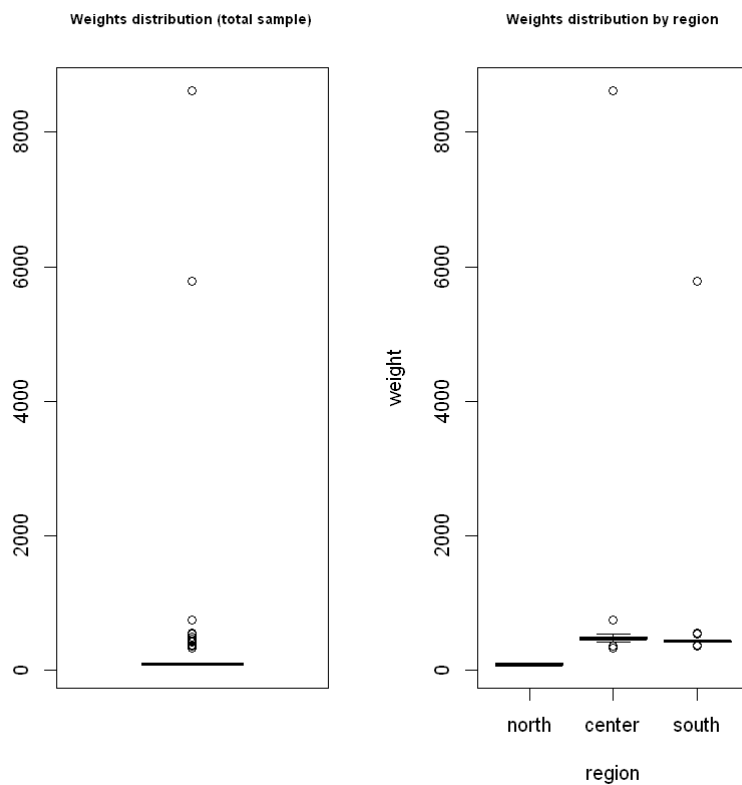
2258507

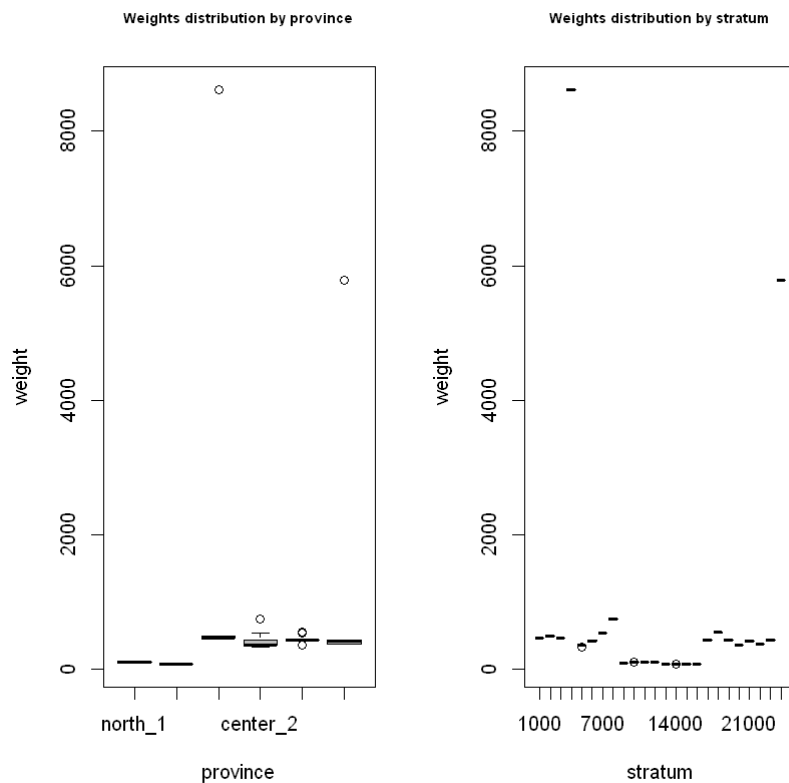
In [58]:

```

## Plot of weights distribution
par(mfrow=c(1, 2))
boxplot(samp$weight,col="grey")
title("Weights distribution (total sample)",cex.main=0.7)
boxplot(weight ~ region, data=samp,col="grey")
title("Weights distribution by region",cex.main=0.7)
par(mfrow=c(1, 2))
boxplot(weight ~ province, data=samp,col="grey")
title("Weights distribution by province",cex.main=0.7)
boxplot(weight ~ stratum, data=samp,col="grey")
title("Weights distribution by stratum",cex.main=0.7)

```





Precision constraints compliance control (by simulation)

In [59]:

```
selected_PSU <- sample_2st[[4]]
df=pop
df$one <- 1
PSU_code="municipality"
SSU_code="id_ind"
PSU_sampled=selected_PSU[selected_PSU$Sampled_PSU==1,]
target_vars <- c("income_hh",
                 "active",
                 "inactive",
                 "unemployed")
PSU_sampled <- selected_PSU[selected_PSU$PSU_final_sample_unit > 0,]
```

In [60]:

```
# Domain level = national
domain_var <- "one"
eval <- eval_2stage(df,
                    PSU_code,
                    SSU_code,
                    domain_var,
                    target_vars,
                    PSU_sampled,
                    nsampl=100,
                    writeFiles=FALSE,
                    progress=TRUE)
```

|=====| 100%

A data.frame: 1 × 5

CV1	CV2	CV3	CV4	dom
<dbl>	<dbl>	<dbl>	<dbl>	<chr>

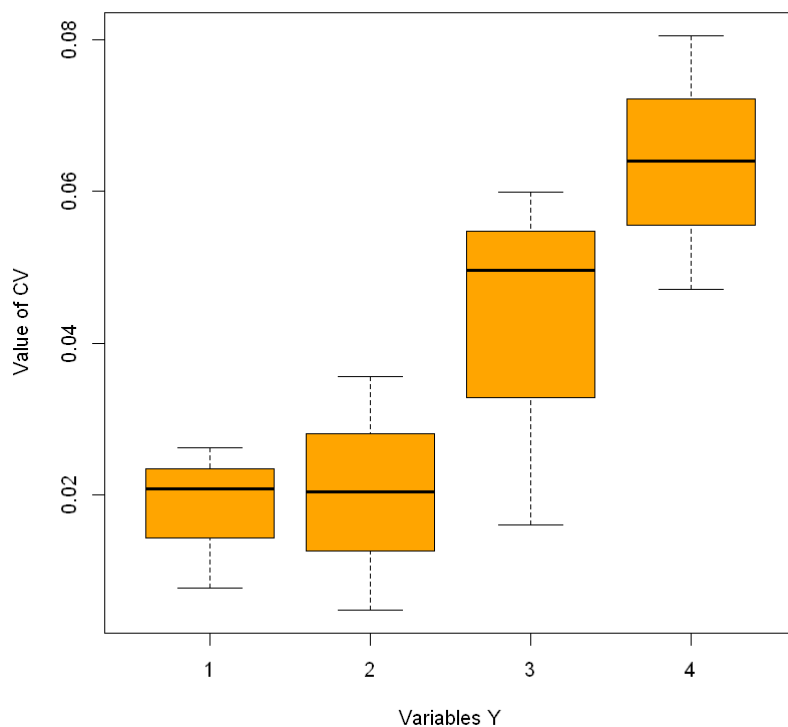
CV1	CV2	CV3	CV4	dom
<dbl>	<dbl>	<dbl>	<dbl>	<chr>
0.0091	0.0094	0.0244	0.0378	DOM1

In [62]:

```
# Domain level = regional
domain_var <- "region"
set.seed(1234)
eval <- eval_2stage(df,
                    PSU_code,
                    SSU_code,
                    domain_var,
                    target_vars,
                    PSU_sampled,
                    nsamp=100,
                    writeFiles=FALSE,
                    progress=TRUE)
eval$coeff_var
```

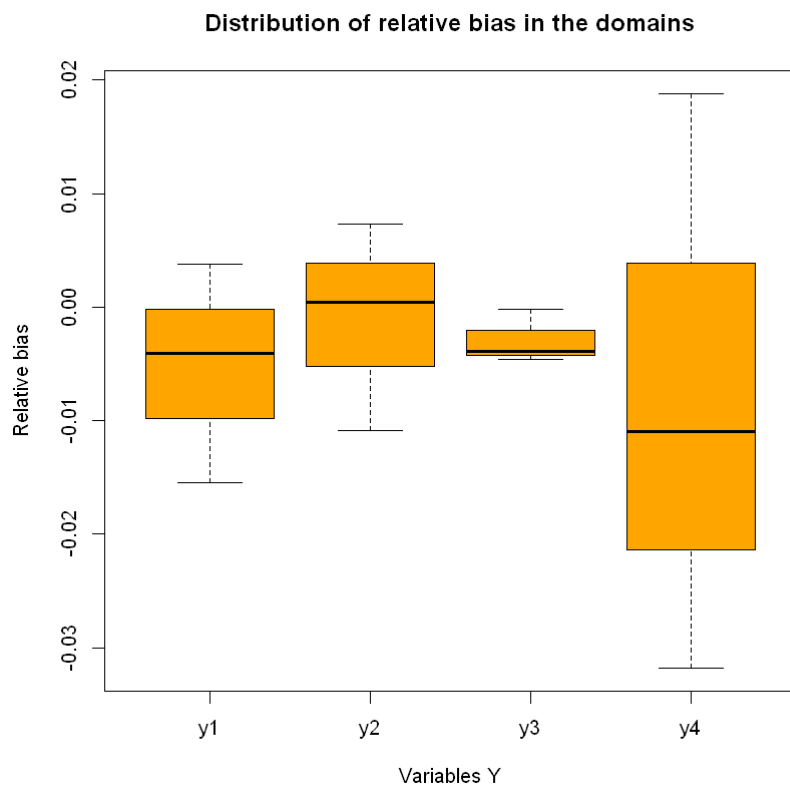
|=====| 100%

Distribution of CV's in the domains



A data.frame: 3 × 5

CV1	CV2	CV3	CV4	dom
<dbl>	<dbl>	<dbl>	<dbl>	<chr>
0.0078	0.0048	0.0160	0.0640	DOM1
0.0209	0.0205	0.0496	0.0805	DOM2
0.0262	0.0356	0.0599	0.0471	DOM3



In [63]: `alloc$sensitivity`

A data.frame: 4 × 6

	Type	Dom	V1	V2	V3	V4
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	DOM1	1	1	1	1	442
5	DOM2	1	1	0	1	2022
9	DOM2	2	1	1	16	123
13	DOM2	3	1	1	1	1

In []: