

# RSSampling

Anna Urbala

11 maja 2020

## RSSampling: A Pioneering Package for Ranked Set Sampling

*by Busra Sevinc, Bekir Cetintav, Melek Esemen, and Selma Gurler*

Link

```
##Loading packages
library("RSSampling")
library("LearnBayes")
## Imperfect ranking example for interested (X) and concomitant (Y) variables
## from multivariate normal dist.
set.seed(1)
mu <- c(10, 8)
variance <- c(5, 3)
a <- matrix(c(1, 0.9, 0.9, 1), 2, 2)
v <- diag(variance)
Sigma <- v%*%a%*%v
x <- rmnorm(10000, mu, Sigma)
xx <- as.numeric(x[,1])
xy <- as.numeric(x[,2])
## Selecting a truncation-based ranked set sample
con.Rrss(xx, xy, m = 4, r = 2, type = "tb", sets = TRUE, concomitant = FALSE, alpha = 0.25)

## $corr.coef
## [1] 0.9040095
##
## $var.of.interest
##      [,1]      [,2]      [,3]      [,4]
## [1,]  1.9048274 11.035541 10.441654 14.28956
## [2,] -2.7455345  5.005396 11.188461 12.69998
## [3,]  8.9901608 12.425687 12.431518 10.76925
## [4,]  7.9035655  8.490473 11.741792 16.08444
## [5,]  5.9616249  5.970233  9.185030 12.16851
## [6,]  0.3453106 10.250466 14.776124 22.55901
## [7,]  5.2518785 13.431264  9.001550 10.50923
## [8,]  3.7943619  7.851410  9.887394 14.29900
##
## $sample.x
##      m = 1      m = 2      m = 3      m = 4
## r = 1 1.904827  5.005396 12.43152 16.08444
## r = 2 5.961625 10.250466  9.00155 14.29900
```

Wartości się różnią mimo ustawionego seeda, do tego nie ma wartości `concomitant.var`.

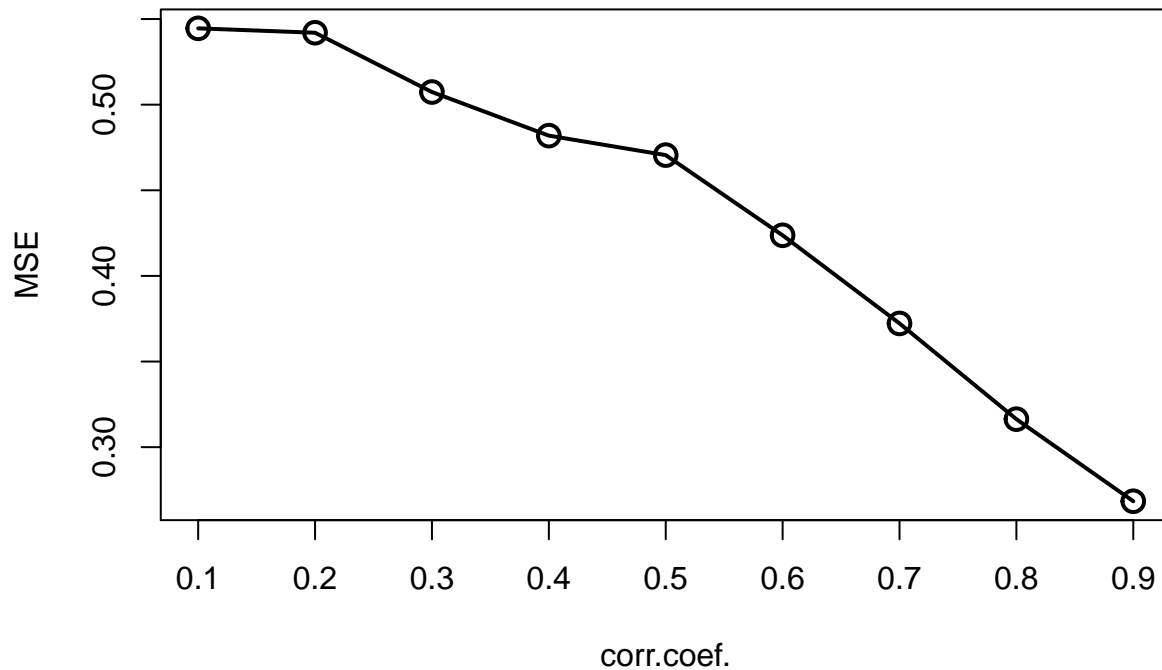
```
## Loading packages
library("RSSampling")
## Generating concomitant variable (Y) from exponential dist.
set.seed(5)
y = rexp(10000)
## Determining the observation numbers of the units which are chosen to sample
obsno.Mrss(y, m = 3, r = 5, type = "m")
```

```
##          m = 1          m = 2          m = 3
## r = 1 "Obs.  5211" "Obs.  4765" "Obs.  3925"
## r = 2 "Obs.  4708" "Obs.  1672" "Obs.  1936"
## r = 3 "Obs.   122" "Obs.  4841" "Obs.  4799"
## r = 4 "Obs.  6905" "Obs.  2320" "Obs.  8194"
## r = 5 "Obs.  4290" "Obs.  6663" "Obs.   910"
```

Znowu inne wartości.

```
library("RSSampling")
library("LearnBayes")
## Imperfect ranking example for interested (X) and concomitant (Y) variables
## from multivariate normal dist.
mu <- c(10, 8)
variance <- c(5, 3)
rho = seq(0, 0.9, 0.1)
se.x = mse.x = numeric()
repeatsize = 1000
for (i in 1:length(rho)) {
  set.seed(1)
  a <- matrix(c(1, rho[i], rho[i], 1), 2, 2)
  v <- diag(variance)
  Sigma <- v%*%a%*%v
  x <- rmnorm(10000, mu, Sigma)
  xx <- as.numeric(x[,1])
  xy <- as.numeric(x[,2])
  for (j in 1:repeatsize) {
    set.seed(j)
    samplex = con.Mrss(xx, xy, m = 5, r = 10, type = "r", sets = FALSE,
                      concomitant = FALSE)$sample.x
    se.x[j] = (mean(samplex) - mu[1])^2
  }
  mse.x[i] = sum(se.x)/repeatsize
}
plot(rho[-1], mse.x[-1], type = "o", lwd = 2,
      main = "MSE values based on increasing correlation levels",
      xlab = "corr.coef.", ylab = "MSE", cex = 1.5, xaxt = "n")
axis(1, at = seq(0.1, 0.9, by = 0.1))
```

## MSE values based on increasing correlation levels



Wykres się różni, prawdopodobnie jest to spowodowane tym co wyżej.

```
abaloneData <- read.csv(url("https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.csv"))
```

```
cor(abaloneData$viscera.weight, abaloneData$whole.weight)
```

```
## [1] 0.9663751
```

ok

```
set.seed(50)
```

```
sampleRSS = con.rss(abaloneData$viscera.weight, abaloneData$whole.weight, m = 5, r = 5, sets = TRUE, conf.level = 0.95)
meanRSS(sampleRSS, m = 5, r = 5, alpha = 0.05, alternative = "two.sided", mu_0 = 0.18)
```

```
## $mean
```

```
## [1] 0.19788
```

```
##
```

```
## $CI
```

```
## [1] 0.1508736 0.2448864
```

```
##
```

```
## $z.test
```

```
## [1] 0.7455195
```

```
##
```

```
## $p.value
```

```
## [1] 0.4559578
```

znowu inne wartości

```
varRSS(sampleRSS, m = 5, r = 5, type = "Stokes")
```

```
## [1] 0.01351398
```

jw.

## Problemy

- inne wartości mimo ustawiania seeda

## Jak naprawić

- nie da się

## Podsumowanie

Kategoria	Ocena
Dostęp do zewnętrznych zasobów	+++++
Kompatybilność z nowszymi wersjami	+++++
Kwestie graficzne/estetyczne	++++x
Brak problemów przy dodatkowej konfiguracji	-----
Odporność na wpływ losowości	xxxxx
Dostępność kodów źródłowych	+++++

### Session info

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 9 (stretch)
##
## Matrix products: default
## BLAS:   /usr/lib/openblas-base/libblas.so.3
## LAPACK: /usr/lib/libopenblas-r0.2.19.so
##
## locale:
##  [1] LC_CTYPE=pl_PL.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=pl_PL.UTF-8      LC_COLLATE=pl_PL.UTF-8
##  [5] LC_MONETARY=pl_PL.UTF-8  LC_MESSAGES=pl_PL.UTF-8
##  [7] LC_PAPER=pl_PL.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=pl_PL.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] LearnBayes_2.15.1 RSSampling_1.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.4.6      codetools_0.2-16 digest_0.6.25    magrittr_1.5
##  [5] evaluate_0.14     icon_0.1.0       rlang_0.4.5     stringi_1.4.6
##  [9] rmarkdown_2.1    tools_3.6.3     stringr_1.4.0   xfun_0.13
## [13] yaml_2.2.1        compiler_3.6.3  htmltools_0.4.0 knitr_1.28
```