Package ‘stratallo’

May 25, 2020

Title Optimum Sample Allocation in Stratified Random Sampling Scheme

Version 0.1.0

Description Functions in this package provide solution to classical problem in survey methodology - an optimum sample allocation in stratified sampling scheme with simple random sampling without replacement design in each stratum.
In this context, the optimal allocation is in the classical Tschuprow-Neyman's sense,
and it satisfies additional upper bounds restrictions imposed on sample sizes in strata.
There are four different algorithms available to use, and one of them is Neyman optimal allocation applied in a recursive way.
All the algorithms are described in detail in ``Wojciak W. Optimal allocation in stratified sampling schemes.
Master's diploma thesis, Warsaw University of Technology. 2019'',
available on-line at <http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf>.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

NeedsCompilation no

Author Wojciech Wojciak [aut, cre],
Jacek Wesolowski [sad],
Robert Wieczorkowski [ctb]

Maintainer Wojciech Wojciak <wojciech.wojciak@gmail.com>

Repository CRAN

Date/Publication 2020-05-25 19:10:06 UTC

R topics documented:

`coma` *Change of Monotonicity Algorithm for Optimal Sample Allocation in Stratified Sampling Scheme*

Description

Implementation of the Change of Monotonicity algorithm for optimal sample allocation problem in stratified sampling scheme with simple random sampling without replacement design in each stratum. The algorithm is described in Wojciak (2019), under the name Sequential Allocation (ver. 2).

Usage

```
coma(d, M, n, verbose = FALSE)
```

Arguments

`d`	a numeric vector, equal to (element-wise) multiplication of strata sizes by standard deviations of a study variable in strata.
`M`	a numeric vector, upper bounds constraints on sample sizes in strata. It must be that `0` < `M` <= `N`, where `N` `-strata sizes`.
`n`	integer, total sample size.
`verbose`	logical, if *TRUE*, more interim variables from chosen algorithm are returned; in particular: `J` - set of strata indices sorted in non-increasing order of `\frac{d}{M}`, `J_` - set of strata indices for which the allocated sample size is strictly less than its upper bound constraint `M`.

Value

Numeric vector with optimal sample allocations in strata, or list with numeric vector of optimal sample allocations in strata (1st element of the list), and corresponding interim variables (further elements of the list).

References

Wojciak W. (2019) *Optimal allocation in stratified sampling schemes*. Master's diploma thesis, Warsaw University of Technology. http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf

Examples

```
N <- c(3000, 4000, 5000, 2000)
S <- rep(1, 4)
coma(d = N * S, M = c(100, 90, 70, 80), n = 190)
```

`Dst` *Variance of Stratified Pi-estimator of the Total*

Description

Compute the variance of the stratified pi-estimator of the population total of the study variable under simple random sampling without replacement design in each stratum. This variance takes the following form
<center>D(n_1,...,n_H) = \sum_{h = 1}^{H} \frac{N_h^2 S_h^2}{n_h} -\sum_{h = 1}^{H} N_h S_h^2,</center>
where H denotes total number of strata, N_1,...,N_H denote strata sizes, and S_1,...,S_H denote standard deviations of a study variable in strata.

Usage

```
Dst(n, N, S)
```

Arguments

n	a numeric vector, sample allocations in strata.
N	a numeric vector, strata sizes.
S	a numeric vector, standard deviations of a study variable in strata.

Value

The value of the variance.

References

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992) *Model Assisted Survey Sampling*. Chapter 3.7 *Stratified Sampling*. Springer.

Examples

```
N <- c(3000, 4000, 5000, 2000)
S <- rep(1, 4)
n_ <- nopt(n = 190, N = N, S = S, M = c(100, 90, 70, 80))
Dst(n = n_, N, S)
```

`nopt` *Optimal Sample Allocation in Stratified Random Sampling Scheme*

Description

A classical problem in survey methodology in stratified sampling schemes is an optimum sample allocation problem. The problem is formulated as the determination of the vector (called sample allocation), that minimizes, under given constraints, the variance of the estimator of the population total of a study variable. For simple random sampling without replacement design in each stratum, this variance takes the following form
<center>$D(n_1,...,n_H) = \sum_{h = 1}^{H} \frac{N_h^2 S_h^2}{n_h} -\sum_{h = 1}^{H} N_h S_h^2$,</center>
where H denotes total number of strata, $N_1,...,N_H$ denote strata sizes, and $S_1,...,S_H$ denote standard deviations of a study variable in strata.

The `nopt` computes the `argmin` $D(n_1,...,n_H)$, under the following constraints <p style="margin-left: 40px">1: $\sum_{h=1}^{H} n_h = n$
2: $n_h \le M_h \le N_h, h = 1,...,H$, </p> where n denotes overall sample size, and $M_1,...,M_H$ denote upper bounds optionally imposed on sample strata sizes.

There are four different underlying algorithms available to use, abbreviated as: *"sa0"*, *"sa1"*, *"coma"*, *"rNa"*. The algorithms are described in detail in Wojciak (2019).

Usage

```
nopt(n, N, S, M = N, method = "sa0", verbose = FALSE, variance = FALSE)
```

Arguments

`n`	integer, total sample size. It must be `n > sum(M)`, otherwise `n` is truncated such that `n = sum(M)`, and therefore allocation returned is equal to `M`. In case of truncation, the warning message is thrown.
`N`	a numeric vector, strata sizes.
`S`	a numeric vector, standard deviations of a study variable in strata.
`M`	a numeric vector, upper bounds constraints on sample sizes in strata. It must be that `0 < M <= N`, where `N -strata sizes`.
`method`	a character string indicating the algorithm to be used. One of *"sa0"* (default), *"sa1"*, *"coma"*, *"rNa"*, can be abbreviated.
`verbose`	logical, if *TRUE*, more interim variables from chosen algorithm are returned.
`variance`	logical, if *TRUE*, the value of variance `D` corresponding to the optimal solution is returned.

Value

Numeric vector with optimal sample allocations in strata, or list with numeric vector of optimal sample allocations in strata (1st element of the list), and corresponding interim variables and/or variance (further elements of the list).

References

Wojciak W. (2019) *Optimal allocation in stratified sampling schemes*. Master's diploma thesis, Warsaw University of Technology. http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992) *Model Assisted Survey Sampling*. Chapter 3.7 *Stratified Sampling*. Springer.

Examples

```
nopt(n = 190, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 40.71 54.29 67.86 27.14
nopt(n = 270, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 66.66 88.88 70.00 44.44
nopt(n = 300, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 84 90 70 56
nopt(n = 330, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 100  90  70  70
nopt(n = 340, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 100  90  70  80

# Example of sample size n truncation, as it exceeds the sum of upper bounds, i.e. n > sum(M).
nopt(n = 350, N = c(3000, 4000, 5000, 2000), S = rep(1, 4), M = c(100, 90, 70, 80))
# 100  90  70  80

N <- pop969[, "N"]
S <- pop969[, "S"]
n <- floor(0.01 * sum(N))
nopt(n, N, S)
nopt(n, N, S, variance = TRUE)

# Example of execution-time comparision of different algorithms using bench R package.
# nfrac <- seq(0.01, 0.9, 0.05)
# n <- setNames(nfrac * sum(N), nfrac)
# lapply(n, function(ni) bench::mark(nopt(ni, N, S, method = "sa0"),
#                                    nopt(ni, N, S, method = "sa1"),
#                                    nopt(ni, N, S, method = "coma"),
#                                    nopt(ni, N, S, method = "rNa"),
#                                    iterations = 200)[c(1,3)])
```

---

pop507 *Example Population with 507 Strata*

Description

Artificial, example population with 507 strata, defined by strata sizes and standard deviations of a variable under study in strata.

Usage

```
pop507
```

Format

A matrix with 507 rows and 2 columns

N stratum sizes

S standard deviations of a variable under study in strata

`pop969` *Example Population with 969 Strata*

Description

Artificial, example population with 969 strata, defined by strata sizes and standard deviations of a variable under study in strata.

Usage

```
pop969
```

Format

A matrix with 969 rows and 2 columns

N stratum sizes

S standard deviations of a variable under study in strata `rNa` *Recursive Neyman Optimal Allocation Algorithm for Optimal Sample Allocation in Stratified Sampling Scheme*

Description

Implementation of the Recursive Neyman optimal allocation algorithm, with enhancement allowing `M <= N` (without enhancement `M = N`), where N denotes vector of strata sizes. The Recursion Neyman optimal allocation algorithm is described in Remark 12.7.1 in Sarndal et al. (1992), and it computes the optimal sample allocation in stratified sampling scheme with simple random sampling without replacement design in each stratum. Details of the enhancement, including the proof are given in Wojciak (2019).

Usage

```
rNa(d, M, n, verbose = FALSE)
```

Arguments

`d`	a numeric vector, equal to (element-wise) multiplication of strata sizes by standard deviations of a study variable in strata.
`M`	a numeric vector, upper bounds constraints on sample sizes in strata. It must be that `0 < M <= N`, where `N -strata sizes`.
`n`	integer, total sample size.
`verbose`	logical, if *TRUE*, more interim variables from chosen algorithm are returned. Not used in current version.

Value

Numeric vector with optimal sample allocations in strata.

References

Wojciak W. (2019) *Optimal allocation in stratified sampling schemes*. Master's diploma thesis, Warsaw University of Technology. http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992) *Model Assisted Survey Sampling*. Springer.

Examples

```
N <- c(3000, 4000, 5000, 2000)
S <- rep(1, 4)
rNa(d = N * S, M = c(100, 90, 70, 80), n = 190)
```

sa0 *Sequential Allocation (ver. 0) Algorithm for Optimal Sample Allocation in Stratified Sampling Scheme*

Description

Implementation of the Sequential Allocation (ver. 0) algorithm for optimal sample allocation problem in stratified sampling scheme with simple random sampling without replacement design in each stratum. The algorithm is described in Wojciak (2019).

Usage

```
sa0(d, M, n, verbose = FALSE)
```

Arguments

`d`	a numeric vector, equal to (element-wise) multiplication of strata sizes by standard deviations of a study variable in strata.
`M`	a numeric vector, upper bounds constraints on sample sizes in strata. It must be that `0` < `M` <= `N`, where `N` `-strata sizes`.
`n`	integer, total sample size.
`verbose`	logical, if *TRUE*, more interim variables from chosen algorithm are returned; in particular: `J` - set of strata indices sorted in non-increasing order of `\frac{d}{M}`, `J_` - set of strata indices for which the allocated sample size is strictly less than its upper bound constraint `M`.

Value

Numeric vector with optimal sample allocations in strata, or list with numeric vector of optimal sample allocations in strata (1st element of the list), and corresponding interim variables (further elements of the list).

References

Wojciak W. (2019) *Optimal allocation in stratified sampling schemes*. Master's diploma thesis, Warsaw University of Technology. http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf

Examples

```
N <- c(3000, 4000, 5000, 2000)
S <- rep(1, 4)
sa0(d = N * S, M = c(100, 90, 70, 80), n = 190)
```

sa1 — *Sequential Allocation (ver. 1) Algorithm for Optimal Sample Allocation in Stratified Sampling Scheme*

Description

Implementation of the Sequential Allocation (ver. 1) algorithm for optimal sample allocation problem in stratified sampling scheme with simple random sampling without replacement design in each stratum. The algorithm is described in Wojciak (2019).

Usage

```
sa1(d, M, n, verbose = FALSE)
```

Arguments

`d`	a numeric vector, equal to (element-wise) multiplication of strata sizes by standard deviations of a study variable in strata.
`M`	a numeric vector, upper bounds constraints on sample sizes in strata. It must be that `0` < `M` <= `N`, where `N` `-strata sizes`.
`n`	integer, total sample size.
`verbose`	logical, if *TRUE*, more interim variables from chosen algorithm are returned; in particular: `J` - set of strata indices sorted in non-increasing order of `\frac{d}{M}`, `J_` - set of strata indices for which the allocated sample size is strictly less than its upper bound constraint `M`.

Value

Numeric vector with optimal sample allocations in strata, or list with numeric vector of optimal sample allocations in strata (1st element of the list), and corresponding interim variables (further elements of the list).

References

Wojciak W. (2019) *Optimal allocation in stratified sampling schemes*. Master's diploma thesis, Warsaw University of Technology. http://home.elka.pw.edu.pl/~wwojciak/msc_optimal_allocation.pdf

Examples

```
N <- c(3000, 4000, 5000, 2000)
S <- rep(1, 4)
sa1(d = N * S, M = c(100, 90, 70, 80), n = 190)
```

Index