Package 'bethel'

July 22, 2025

Version 0.2

Date 2009-05-19

Author Michele De Meo

Maintainer Michele De Meo <micheledemeo@gmail.com>

Title Bethel's algorithm.

Description The sample size according to the Bethel's procedure.

Depends R (>= 2.5.0)

License GPL (>= 2)

Repository CRAN

Date/Publication 2012-10-29 08:58:17

NeedsCompilation no

Contents

---

bth	*The Bethel algorithm*

Description

Bethel procedure (1989) allows to determine total sample size and allocation of units in strata, so to minimize costs under the constraints of defined precision levels of estimates (coefficient of variation: CV), in the multivariate case (more than one estimate).Input to this algorithm is given by the information on distributional characteristics (total and variance) of target variables in the population strata.

Usage

```
bth(S, T, eps = 1e-10)
```

Arguments

S	A dataframe or a matrix with strata, variances, population size and minumum sample size. See details below.
T	A dataframe or a matrix with precision levels (coefficient of variation: CV) and totals. See details below.
eps	The level of precision for the algorithm convergence. The default is 1e-10.

Details

The Bethel algorithm allows the calculation of the sample size (in each strata) in the case of a multi-variate and stratified population. The input of the procedure consists of two dataframe, respectively, S and T.

S is composed by a minimum of 6 columns (ncol(S) > = 6), suppose ncol(S) = k. The first column shows the strata labels. The k-th column shows the minimum sample rate for each strata, such as 0.04 if the sample will consist of at least 4% in each strata. Similarly, the (k-1)-th column contains the absolute minimum sample size (for example the value 3 if the each strata has to be composed at least of 3 sample units). The (k-2)-th column shows the unit cost per interview(in each strata). Generally this value is equal to 1 to indicate the same cost in all strata. The (k-3)-th column gives the size of the population in each strata. Finally, the estimated variances for the k-5 observed variables are shown in columns from the second to (k-6)-th. See the example below.

T is composed of 2 columns. The first column shows the coefficients of variation (CV) for k-6 variables analyzed (for example, CV = 0.05 for each variable). The second column shows the estimated totals for the same k-6 variables. See the example below.

Value

B	The dataframe with the Bethel sample size (bethelNum) and the minimum sample size (bethelNum2). If the sample size (in a generic strata) according the Bethel algorithm is equal to 2, then bethelNum will be equal to 2, but bethelNum2 will be 3 if, for example, 3 is the minimum sample size specified in the (k-1)-th column of S. See the example below.

Author(s)

Michele De Meo <micheledemeo@gmail.com>

References

Bethel, J.W. (1989), *Sample Allocation in Multivariate Surveys*. Survey Methodology, Vol. 15, pp. 47-57.
Chromy, J. B. (1987), *Design Optimization With Multiple Objectives*. Proceedings of the Section on Survey Research Methods, 1987. American Statistical Association, pp. 194-199.

See Also

tapply, var

Examples

```
#Given a population of 1000 individuals (dataframe pop)
#classified according to sex and geographic area, we have collected
#yarly data on the following variables: income, number of books read,
#total days of sporting activities. To run a survey and to obtain
#the total estimates of these 3 variables (total income,total number
#of book, total number of days) we calculate the sample size to obtain,
#for example, a precision level (coefficient of variation) of 0.05.

library(bethel)
data(pop)
attach(pop)
str(pop)

#Calculate the dataframe with:
##- strata labels
##- estimated variances
##- number of population units

b1<-as.data.frame(cbind(var_Income=tapply(income,strata,var),
var_books=tapply(books,strata,var),
var_days=tapply(sportDays,strata,var),
num_units=tapply(sportDays,strata,length)))
b1<-cbind(strata=row.names(b1),b1)
row.names(b1)<-NULL

#Add 3 columns:
##- unit cost per interview
##- minimum sample size n/N (where N is the population size)
##- minimum sample size n

b1<-cbind(b1, c=rep(1,8), n=rep(3,8), n_2=rep(0.04,8))

#Calculate dataframe with:
##- precision levels (coefficients of variation)
##- total estimates

b2<-as.data.frame(cbind(CV=rep(0.05,3), tot=colSums(pop[,2:4])))

#Bethel sample according to a precision level (CV) of 0.05

bth(b1,b2)

#Bethel sample according to different precision level (CV)

b2<-as.data.frame(cbind(CV=c(0.05,0.01,0.2), tot=colSums(pop[,2:4])))
bth(b1,b2)
```

---

pop   *Bethel population*

Description

1000 individuals classified according to sex (M,F) and geographical area (area1 to area4). Collected variables: yearly data on income, number of books read, total days of sporting activities.

Usage

```
data(pop)
```

Format

A data frame with 1000 observations on the following 4 variables.

strata a factor with levels F_area1 F_area2 F_area3 F_area4 M_area1 M_area2 M_area3 M_area4

income yarly income

books number of books read

sportDays total days of sporting activities

Examples

```
data(pop)
str(pop)
summary(pop)
```

Index