# R2BEAT: AN R PACKAGE FOR THE OPTIMAL SAMPLE ALLOCATION



Conferenza Nazionale di Statistica

GIULIO BARCAROLI-Independent consultant in statistical survey methodologies gbarcaroli@gmail.com | ILARIA BOMBELLI-Istat ilaria.bombelli@istat.it |

ANDREA FASULO-EUROSTAT and rea. fasulo@ec.europa.eu | ALESSIO GUANDALINI-Istat alessio.guandalini@istat.it | MARCO D. TERRIBILI-Istat terribili@istat.it |

### Abstract

R2BEAT ("R 'to' Bethel Extended Allocation for Two-stage sampling") is an R package for the optimal allocation of a sample. Its peculiarity lies in properly addressing **allocation problems** for **two-stage and complex sampling designs** with **multi-domain and multi-purpose aims**. This is common in many official and non-official statistical surveys, therefore R2BEAT could become an essential tool for planning a sample survey. The functions implemented in R2BEAT allow the use of different workflows, depending on the available information on one or more interest variables. The package covers all the phases, from the optimization of the sample to the selection of the Primary and Secondary Stage Units. Furthermore, it provides several outputs for evaluating the allocation results.

## **Main functions**

prepareInputToAllocation\_beat.1st(): returns a dataframe, starting from the sampling frame (either universe or sample of a previous survey) with strata information.

prepareInputToAllocation1(): returns Input dataframes for R2BEAT two-stages sample design when sampling frame is available.

beat.1st(): Compute one stage multivariate optimal allocation.

beat.2st(): Multivariate optimal allocation for different domains in two stage statified sample design.

select\_PSU(): Select sample of primary stage units (PSU).

select\_SSU(): Select sample of secondary stage units (SSU).

**beat.1cv()**: The function returns a dataframe with planned and actual coefficients of variation (CV) in a multivariate multi-domain allocation problem.

More functions available at <a href="https://barcaroli.github.io/R2BEAT/reference/index">https://barcaroli.github.io/R2BEAT/reference/index</a>

## Algorithm for two-stage sampling design

#### Input:

- precision constraints in terms of CV;
- information on sampling strata (mean and stdev of target variables, N, ...);



- information on previous design: deff, effst, ρ;
- information on PSUs in sampling strata (measure of size);
- minimum number of SSUs per PSU.

#### Output:

- for each stratum: number of PSUs and SSUs to be selected;
- expected CVs for target estimates;
- sensitivity of the solution.



## **Comparison with competitors**



## Conclusions

**Methodology**: is based on the Bethel algorithms and the resulting sample designs are compliant with the precision constraints set on the target estimates of a given survey.

#### **Completeness**:

a. In terms of applicability: R2BEAT applies to stratified,

R2BEAT always provide samples with smaller sample sizes, both in terms of PSUs and SSUs.

## Main references and useful links

multipurpose and multidomain, onestage and two-stage sampling surveys;

b. In terms of coverage of the user needs: the package covers all the steps of a complex sample design, from the setting of precision constraints, the determination of the total sample size and of the allocation in the strata and the selection of the sampling units, distinctly according to the stage of selection.

**Efficiency:** In our experiments, on equal errors, the sample size determined by R2BEAT is always lower, than the ones determined by two other competitor packages in terms of both Primary and Secondary Stage Units (PSUs and SSUs).

Barcaroli, et al., "Two-Stage Sampling Design and Sample Selection with the R Package R2BEAT", *The R Journal*, 2023 Bethel, J. (1989) Sample allocation in multivariate surveys. *Survey methodology*, 15.1: 47-57 Cochran, W. (1977) *Sampling Techniques*. John Wiley & Sons, Inc., New York Link to Github repository: <u>https://barcaroli.github.io/R2BEAT/index</u>