

Streszczenie

Alokacja optymalna próby w warstwowym schemacie próbkowania jest jednym z podstawowych zagadnień metodologii badań reprezentacyjnych. W klasycznej postaci zagadnienie to polega na wyznaczeniu liczebności prób w warstwach przy ustalonej liczebności całkowitej próby, z uwzględnieniem kryteriów optymalności związanych z minimalizacją wariancji estymatorów.

Niniejsza praca koncentruje się na problemie wielodomenowej alokacji optymalnej próby w schemacie warstwowym. Celem jest jednoczesna minimalizacja wariancji estymatora sumy globalnej oraz wariancji estymatorów sum w poszczególnych domenach, przy ustalonej całkowitej liczebności próby i zapewnieniu, że przydziały nie przekraczają liczebności populacji w warstwach. Problem ten sformułowano w sposób umożliwiając kontrolę względnej precyzji estymatorów w poszczególnych domenach za pomocą uprzednio określonych wag priorytetowych. Z dostępnej literatury wynika, że dotychczas nie opracowano dokładnego algorytmu analitycznego, który rozwiązywałby ten problem.

Główne rezultaty pracy obejmują sformułowanie warunków dostatecznych dla rozwiązania optymalnego rozważanego problemu alokacji oraz zdefiniowanie algorytmu RDCA, który rozwiązuje ten problem w oparciu o ustanowione warunki optymalności. Warunki optymalności wyprowadzono przy użyciu metod i narzędzi optymalizacji wypukłej (w tym warunków Karusha-Kuhna-Tuckera) oraz algebry macierzy w kontekście zagadnienia ich wartości własnych. Dowód poprawności zaproponowanego algorytmu stanowi istotny element pracy.

Integralną częścią dysertacji jest implementacja algorytmu RDCA jako funkcji w języku R, zawartej w pakiecie **stratallo**, dostępnym publicznie w repozytorium Comprehensive R Archive Network.

Słowa kluczowe: alokacja optymalna próby, alokacja wielodomenowa, minimalizacja wariancji, kontrolowana precyzja, próbkowanie warstwowe, dokładny algorytm, algorytm RDCA, warunki Karusha-Kuhna-Tuckera, wartości własne, wektory własne.

Abstract

Optimum sample allocation in stratified sampling is one of the fundamental issues in survey methodology. In its classical formulation, it involves determining how a given total sample size should be allocated among strata so that certain criteria, typically related to minimizing the variances of estimators, are satisfied.

This thesis addresses the multi-domain optimum sample allocation problem in stratified sampling. The objective is the simultaneous minimization of the variance of the global total estimator and the variances of domain total estimators under a fixed total sample size, while ensuring that allocations do not exceed the population sizes within strata. The problem has been formulated in a way that allows controlling the relative precision of estimators across domains through pre-specified domain-wise priority weights. A review of the available literature indicates that no exact analytical algorithm has yet been developed to solve this problem.

The main contributions of this thesis are twofold. First, it formulates sufficient optimality conditions for the solution to the allocation problem under consideration. Second, it defines the RDCA algorithm, which solves this allocation problem based on the established conditions. The optimality conditions were derived using convex optimization techniques (e.g., Karush-Kuhn-Tucker conditions) and matrix algebra methods, with analyses involving the eigenvalue problem. A proof of correctness of the proposed algorithm constitutes an essential part of the thesis.

An integral part of this dissertation is the implementation of the RDCA algorithm as a function in the R programming language, provided in the **stratallo** package, which is publicly available in the Comprehensive R Archive Network.

Keywords: optimum sample allocation, multi-domain allocation, variance minimization, controlled-precision, stratified sampling, exact algorithm, RDCA algorithm, Karush-Kuhn-Tucker conditions, eigenvalues, eigenvectors.