VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
INSTITUTE OF MATHEMATICS AND INFORMATICS

Dalius PUMPUTIS

STATISTICAL ESTIMATORS OF THE FINITE POPULATION PARAMETERS IN THE PRESENCE OF AUXILIARY INFORMATION

Summary of Doctoral Dissertation
Physical Sciences, Mathematics (01P)

Vilnius VGTU LEIDYKLA TECHNIKA 2008

Doctoral dissertation was prepared at the Institute of Mathematics and Informatics in 2004–2008.

Scientific Supervisor

Prof Dr Habil Romanas JANUŠKEVIČIUS (Vilnius Pedagogical University, Physical Sciences, Mathematics – 01P).

Consultant

Assoc Prof Dr Aleksandras PLIKUSAS (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P).

The dissertation is being defended at the Council of Scientific Field of Mathematics at Vilnius Gediminas Technical University:

Chairman

Prof Dr Habil Kęstutis KUBILIUS (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P).

Members:

Dr Habil Vidmantas BENTKUS (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P),

Prof Dr Habil Algimantas BIKELIS (Vytautas Magnus University, Physical Sciences, Mathematics – 01P),

Prof Dr Habil Remigijus LEIPUS (Vilnius University, Physical Sciences, Mathematics – 01P),

Prof Dr Habil Leonas SAULIS (Vilnius Gediminas Technical University, Physical Sciences, Mathematics – 01P).

Opponents:

Prof Dr Habil Eugenijus MANSTAVIČIUS (Vilnius University, Physical Sciences, Mathematics – 01P),

Assoc Prof Dr Marijus RADAVIČIUS (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P).

The dissertation will be defended at the public meeting of the Council of Scientific Field of Mathematics in the Conference and Seminars Center of the Institute of Mathematics and Informatics at 2 p. m. on 9 February 2009.

Address: A. Goštauto g. 12, LT-01108 Vilnius, Lithuania.
Tel. +370 5 274 4952, +370 5 274 4956; fax +370 5 270 0112;
e-mail: doktor@adm.vgtu.lt

The summary of the doctoral dissertation was distributed on 8 January 2009.

A copy of the doctoral dissertation is available for review at the Libraries of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania) and the Institute of Mathematics and Informatics (Akademijos 4, LT-08663 Vilnius, Lithuania).

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

Dalius PUMPUTIS

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI, GAUTI NAUDOJANT PAPILDOMĄ INFORMACIJĄ

Daktaro disertacijos santrauka
Fiziniai mokslai, matematika (01P)

Vilnius VGTU LEIDYKLA TECHNIKA 2008

Disertacija rengta 2004–2008 metais Matematikos ir informatikos institute.

Mokslinis vadovas

prof. habil. dr. Romanas JANUŠKEVIČIUS (Vilniaus pedagoginis universitetas, fiziniai mokslai, matematika – 01P).

Konsultantas

doc. dr. Aleksandras PLIKUSAS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).

Disertacija ginama Vilniaus Gedimino technikos universiteto Matematikos mokslo krypties taryboje:

Pirmininkas

prof. habil. dr. Kęstutis KUBILIUS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).

Nariai:

habil. dr. Vidmantas BENTKUS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P),

prof. habil. dr. Algimantas BIKELIS (Vytauto Didžiojo universitetas, fiziniai mokslai, matematika – 01P),

prof. habil. dr. Remigijus LEIPUS (Vilniaus universitetas, fiziniai mokslai, matematika – 01P),

prof. habil. dr. Leonas SAULIS (Vilniaus Gedimino technikos universitetas, fiziniai mokslai, matematika – 01P).

Oponentai:

prof. habil. dr. Eugenijus MANSTAVIČIUS (Vilniaus universitetas, fiziniai mokslai, matematika – 01P),

doc. dr. Marijus RADAVIČIUS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).

Disertacija bus ginama viešame Matematikos mokslo krypties tarybos posėdyje 2009 m. vasario 9 d. 14 val. Matematikos ir informatikos instituto konferencijų ir seminarų centre.

Adresas: A. Goštauto g. 12, LT-01108 Vilnius, Lietuva.
Tel.: (8 5) 274 4952, (8 5) 274 4956; faksas (8 5) 270 0112;
el. paštas doktor@adm.vgtu.lt

Disertacijos santrauka išsiuntinėta 2009 m. sausio 8 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva) ir Matematikos ir informatikos instituto (Akademijos g. 4, LT-08663 Vilnius, Lietuva) bibliotekose.

VGTU leidyklos „Technika“ 1577-M mokslo literatūros knyga.

General Characteristic of the Dissertation

Scientific problem

Various sources of auxiliary information are available in nowadays official statistics and other areas. It may be data of various statistical registers and other administrative sources. The dissertation analyzes how to incorporate auxiliary information into the estimation of the finite population parameters, such as finite population total, variance, covariance, and how to use it for the stratification of finite populations.

Topicality of the work

Survey sampling is a young branch of statistics, that has been rapidly developing since 1940. The Soviet Union did not develop this branch of science, therefore the first solid scientific works, written in this field by Lithuanian statisticians, were published only after restitution of independence. Nowadays, the methods of survey sampling are widely applied in the official statistics, social, economic, and other surveys. Therefore it is important to develop this branch of statistics and to improve existing methods of survey sampling.

In many scientific publications, the importance of use of the known auxiliary information is emphasized for constructing estimators of finite population parameters. If auxiliary variables are well correlated with the study variables, then it is possible to obtain more accurate estimates of parameters. The publications of J.C. Deville and C.E. Särndal are particularly significant. Using auxiliary variables, they have introduced a new class of estimators for finite population totals, called as *calibrated estimators*. This type of estimator is increasingly applied in the official statistics.

Except the population total, there exists a lot of other important, but complicated parameters: the ratio of two population totals, population variance, covariance, quantiles and others. The estimators of the ratio of two population totals may be used in the salary surveys, whereas the estimators of population covariance may be applied for estimation of regression coefficients or covariance matrixes. Unfortunately, the estimation of these parameters, using auxiliary variables, is not widely studied in the literature. So, it is important to extend the class of estimators of parameters mentioned above. For this purpose, we are developing the technique of calibration of design weights, which was proposed by J.C. Deville and C.E. Särndal.

The accuracy of estimates and their variance depends not only on the estimators or auxiliary information, but also on the sampling design. If we have some auxiliary information on the structure of population, the stratified sampling design is often effective and widely used in many surveys. Seeking for better survey results, we should use more effective stratification rules, which may be obtained either by improving existing methods or introducing new ones.

Research object

The research object of the work is as follows:

- estimation of the finite population parameters using auxiliary information;
- model-assisted and calibrated estimators of the finite population covariance;
- stratification of finite populations.

The aim and problems of the dissertation

The aim of this dissertation is to improve some methods of stratification and estimation of the finite population total and covariance, using auxiliary information. Let us state the following problems:

1. To construct calibrated estimators of the finite population total, variance, and covariance, using different distance functions and calibration equations.
2. To construct calibrated estimators of the finite population covariance that use several systems of weights.
3. To construct estimators for variance of the calibrated estimators obtained.
4. To compare by simulation the constructed calibrated estimators with the standard estimators of respective parameters.
5. To compare by simulation the constructed calibrated estimators of population covariance (see [A1]) with the linear regression model-assisted and calibrated estimator, proposed by C. Wu and R.R. Sitter.
6. To modify the linear regression model-assisted and calibrated estimator of population covariance, using separate auxiliary variables for each study variable.
7. Assuming that population distribution is exponential, to modify the geometric stratification method, proposed by P. Gunning and J.M. Horgan.

Research methods

The analytic, probabilistic, and experimental methods were applied in the dissertation. The Lagrange multiplier method, Taylor linearization technique, the methods for calculating numerical characteristics of random variables and matrix differentiation rules were used to prove the propositions formulated by the dissertation author. For definition of the adjusted model-calibrated estimator of finite population covariance, the theory of model-calibrated estimators was used. The population stratification approaches, such as the cumulative root frequency rule, geometric and power methods, were reviewed to consider the problem of optimal stratification. The adjusted geometric stratification method, proposed by us, is based on the population distribution. The mathematical computing software Matlab was used to perform all the simulations, described in this dissertation.

Scientific novelty

In this dissertation, we construct calibrated estimators of the finite population total, using C.E. Särndal's idea of weight calibration and several distance functions. The estimators are compared by simulation. Using distance functions that contain, for example, square roots, we can assure that calibrated weights be positive. The negative weights are often the reason for greater variance of estimators. An approximate variance of several constructed estimators is derived in this work as well.

The novelty and originality of this dissertation is that we propose and analyze here the calibrated estimators for finite population covariance (variance), using one or more weighting systems. For definition of the calibrated weights, we introduce new calibration equations, adapted to the estimation of population covariance. The problem of estimating the variance for the constructed estimators of covariance is considered in this dissertation as well. This is a technically complicated task, because an explicit solution of calibration equations does not exist in many cases.

An adjusted linear regression model-assisted and calibrated estimator of population covariance is also treated as a new result. Its main difference from C. Wu and R.R. Sitter's estimator is in the construction, as we use separate auxiliary variables for each study variable.

In this dissertation, we consider the problem of stratification of skewed populations and propose *an adjusted geometric stratification method.*

The new Matlab functions were created by the author to perform all the simulations, described in this dissertation.

Defended propositions

1. The proposition that provides the expressions for calibrated weights in the case of estimation of the finite population total.

2. The propositions that provide the expressions for calibrated weights in the case of estimation of finite population variance and covariance.

3. The calibrated estimators of finite population covariance using several systems of weights.

4. The propositions on the calculation and estimation of approximate variance for the constructed calibrated estimators of population total and covariance (variance).

5. The analysis of influence of different calibration equations on the estimation accuracy.
6. An adjusted geometric stratification method for skewed populations.
7. An adjusted linear regression model-assisted and calibrated estimator of the finite population covariance.

The scope of the scientific work

The dissertation consists of introduction, three chapters and conclusions. In addition, the conceptual dictionary, index and lists of notation and references are added. The total scope of the dissertation – 134 pages, 6 pictures, 12 tables and 193 mathematical expressions. The work cites 116 references.

1. Estimators using Auxiliary Information

In this chapter, we give a wide description of the dissertation topics and review of the main results, obtained by other researchers. The most important results are related to the calibrated estimators of the population total. Let us define them.

Consider a finite population $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$ of N elements. Let y be a study variable defined on the population $\mathcal{U}$ and taking real values $y_1, y_2, \ldots, y_N$. A probability sample $\mathbf{s}$ of size n is drawn from the population $\mathcal{U}$ with a given sampling design such that the inclusion probabilities $\pi_k = P(k \in \mathbf{s})$ and $\pi_{kl} = P(k\&l \in \mathbf{s})$ are strictly positive. Suppose that for each sample element k, the vector of values $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ of J auxiliary variables is known. We assume that the total $\mathbf{t_x} = \sum_{k \in U} \mathbf{x}_k$ is also known. The total $\mathbf{t_x}$ is used for constructing the population total $t_y = \sum_{k=1}^{N} y_k$ estimators called as *calibrated estimators*. J.C. Deville and C.E. Särndal suggested the idea of weight calibration by modifying the weights $d_k = 1/\pi_k$ of the Horvitz-Thompson estimator $\hat{t}_{y\pi} = \sum_{k \in \mathbf{s}} y_k/\pi_k = \sum_{k \in \mathbf{s}} d_k y_k$. The calibrated estimator

$$\hat{t}_{yw} = \sum_{k \in \mathbf{s}} w_k y_k$$

of the total t_y is defined under the following conditions:

1. Using the weights w_k, the known total $\mathbf{t_x}$ is estimated without error:

$$\hat{\mathbf{t}}_{\mathbf{x}} = \sum_{k \in \mathbf{s}} w_k \mathbf{x}_k = \mathbf{t_x};$$

2. The distance between the design weights d_k and calibrated weights w_k is

minimal according to the distance function

$$L(w_k, d_k, k \in \mathbf{s}) = \sum_{k \in \mathbf{s}} G_k(w_k, d_k)/q_k.$$

Here q_k are free additional weights. The function $G_k(w, d)$ satisfies the following conditions:

- for every fixed $d > 0$, $G_k(w, d)$ is nonnegative, differentiable with respect to w, strictly convex, and such that $G_k(d, d) = 0$;
- $g_k(w, d) = \partial G_k(w, d)/\partial w$ is a continuous, strictly increasing function and $g_k(d, d) = 0$.

2. Stratification of Finite Population

In this chapter, we consider the problem of efficient stratification in the case of skewed population, where the mean μ_y of the survey variable y is estimated and an auxiliary variable x is treated as a stratification variable.

Consider a finite population $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$. Let $y : y_1, \ldots, y_N$ be a study variable defined on the population $\mathcal{U}$. Suppose we have an auxiliary variable $x : x_1, \ldots, x_N$. Let us divide the population $\mathcal{U}$ into H strata, where H is a fixed known number. Denote by U_h the stratum h, by $\mathbf{s}, \mathbf{s} \subset \mathcal{U}$, a stratified random sample set, drawn from the population $\mathcal{U}$, and by $\mathbf{s}_h$ a simple random sample selected from the stratum U_h.

The classical stratification problem is formulated by choosing the population mean μ_y as a parameter of interest and minimizing the variance of its estimator:

$$\widehat{\mu}_y = \frac{1}{N} \sum_{h=1}^{H} N_h \bar{y}_h.$$

Here $\bar{y}_h$ is the sample mean in the stratum U_h, N_h is the number of elements in the stratum U_h, and the product $N_h \bar{y}_h$ is a well known Horvitz-Thompson estimator of the stratum U_h total.

We suppose the number of strata H and the sample size n to be chosen, and the sample is distributed according to the Neyman optimal allocation.

Let the variable y be known and its values be arranged in an ascending order. Denote by k_0 and k_H the smallest and largest values of y, respectively. The problem is to find intermediate values $k_1, k_2, \ldots, k_{H-1}$ such that $Var(\widehat{\mu_y})$ be minimal. The values $k_1, k_2, \ldots, k_{H-1}$ are called as *stratum boundaries*. An assumption that the variable y is known is unrealistic, therefore we will use the auxiliary variable x for stratification. The auxiliary variable x should be well correlated with the study variable y. The principle remains the same: the values of variable x are arranged in an ascending order and we are looking for the stratum boundaries which minimize the variance of the mean estimator $Var(\widehat{\mu_x})$ for the variable x.

T. Dalenius has showed that stratum boundaries with the above-mentioned property exist and satisfy the complicated iterative equations. It is difficult to apply them in practise. That encouraged us to analyse the approximation to the exact solution of these equations.

Adjusted geometric method. Using the idea of P. Gunning and J.M. Horgan to equalize the coefficients of variation of each stratum and assuming that the distribution of a stratification variable is exponential, we get iterative equations for defining the approximation to the optimum strata boundaries:

$$k_h^{(adj)} = \frac{I_1(h)I_2(h+1)k_{h+1}^{(adj)} + I_1(h+1)I_2(h)k_{h-1}^{(adj)}}{I_1(h)I_2(h+1) + I_1(h+1)I_2(h)},$$

where

$$I_1(h) = \int_{k_{h-1}^{(adj)}}^{k_h^{(adj)}} te^{\lambda t}dt, \;\; I_2(h) = \int_{k_{h-1}^{(adj)}}^{k_h^{(adj)}} e^{\lambda t}dt.$$

The simulation results show that the adjusted geometric method outperforms the cumulative root frequency method, geometric method, and power method in the considered populations, whose the coefficient of skewness ac satisfies the inequality $ac > 10$.

3. Calibrated Estimators of the Finite Population Total and Covariance

3.1. Calibrated estimators of the finite population total under different distance functions

Using distance function

$$L_1 = \sum_{k\in\mathbf{s}} \frac{(w_k - d_k)^2}{d_k q_k}, \tag{1}$$

J.C. Deville and C.E. Särndal constructed a respective calibrated estimator of the total. We have derived the expressions of the calibrated weights, using another distance functions:

$$L_2 = \sum_{k\in\mathbf{s}} \frac{w_k}{q_k}\log\frac{w_k}{d_k} - \frac{1}{q_k}(w_k - d_k), \qquad L_3 = \sum_{k\in\mathbf{s}} 2\frac{(\sqrt{w_k} - \sqrt{d_k})^2}{q_k},$$

$$L_4 = \sum_{k\in\mathbf{s}} -\frac{d_k}{q_k}\log\frac{w_k}{d_k} + \frac{1}{q_k}(w_k - d_k), \qquad L_5 = \sum_{k\in\mathbf{s}} \frac{(w_k - d_k)^2}{w_k q_k}, \quad (2)$$

$$L_6 = \sum_{k\in\mathbf{s}} \frac{1}{q_k}\left(\frac{w_k}{d_k} - 1\right)^2, \qquad L_7 = \sum_{k\in\mathbf{s}} \frac{1}{q_k}\left(\frac{\sqrt{w_k}}{\sqrt{d_k}} - 1\right)^2.$$

The approximate variance of calibrated estimators may be derived by the Taylor linearization technique, but not always it is used, because often we do not have explicit expressions for the calibrated weights w_k. An explicit solution to the calibration problem exists only for the distance functions L_1 and L_6. In these cases, we propose the expressions for approximate variance of respective estimators.

3.2. Calibrated estimators of the finite population covariance

In this section, we define some calibrated estimators of the population covariance. They employ one or several systems of calibrated weights. Different calibration equations and distance functions are used for definition of weights.

Estimators of the population covariance using one system of weights

Consider a finite population $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$, of N elements. Let y and z be two study variables, defined on the population $\mathcal{U}$ and taking values $\{y_1, y_2, \ldots, y_N\}$ and $\{z_1, z_2, \ldots, z_N\}$, respectively. The values of the variables y and z are not known. We are interested in the estimation of the finite population covariance

$$Cov(y, z) = \frac{1}{N-1}\sum_{k=1}^{N}\big(y_k - \mu_y\big)\big(z_k - \mu_z\big),$$

where

$$\mu_y = \frac{1}{N}\sum_{k=1}^{N} y_k, \quad \mu_z = \frac{1}{N}\sum_{k=1}^{N} z_k.$$

Assume that two auxiliary variables a and b with the population values $\{a_1, a_2, \ldots, a_N\}$ and $\{b_1, b_2, \ldots, b_N\}$ are available. It means that we have an auxiliary vector $\mathbf{a}_k = (a_k, b_k)'$ for every population element k. Let the variable a be as an auxiliary variable for the study variable y and variable b be auxiliary for z. Denote by $Cov(a, b)$ their known covariance. We construct a new calibrated estimator $Cov(y, z)$, using these known auxiliary variables.

We consider the calibrated estimator of the covariance $\widehat{Cov}_w(y, z)$, of the

form
$$\widehat{Cov}_w(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(y_k - \widehat{\mu}_{yw})(z_k - \widehat{\mu}_{zw}), \tag{3}$$

where
$$\widehat{\mu}_{yw} = \frac{1}{N}\sum_{k\in\mathbf{s}} w_k y_k, \quad \widehat{\mu}_{zw} = \frac{1}{N}\sum_{k\in\mathbf{s}} w_k z_k.$$

The calibrated weights w_k are defined similarly as they have been defined when constructing the calibrated estimator of the total:

a) the weights w_k satisfy some calibration equations;

b) the distance between the weights d_k and w_k is minimal according to some distance function L.

We examine three groups of calibrated estimators of the population covariance defined by three calibration equations. Four distance functions are taken for each calibration equation (condition b)).

I. Nonlinear Calibration. Let the calibration equation be

$$\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(a_k - \widehat{\mu}_{aw})(b_k - \widehat{\mu}_{bw}) = Cov(a,b), \tag{4}$$

where
$$\widehat{\mu}_{aw} = \frac{1}{N}\sum_{k\in\mathbf{s}} w_k a_k, \quad \widehat{\mu}_{bw} = \frac{1}{N}\sum_{k\in\mathbf{s}} w_k b_k.$$

We call this case a *nonlinear calibration*, since the calibration equation (4) is nonlinear with respect to the weights w_k.

II. Linear Calibration. The second calibration equation is as follows:

$$\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(a_k - \mu_a)(b_k - \mu_b) = Cov(a,b), \tag{5}$$

$$\mu_a = \frac{1}{N}\sum_{k=1}^{N} a_k, \quad \mu_b = \frac{1}{N}\sum_{k=1}^{N} b_k.$$

We refer to this case as to a *linear calibration*, since we are actually calibrating the total of the variable $(a-\mu_a)(b-\mu_b)$.

III. Calibration of Totals. The system of weights of the third type is defined by the calibration of totals of the auxiliary variables; in this case, the calibration equations are:

$$\sum_{k\in\mathbf{s}} w_k a_k = \sum_{k=1}^{N} a_k, \quad \sum_{k\in\mathbf{s}} w_k b_k = \sum_{k=1}^{N} b_k. \tag{6}$$

This case is motivated by the current survey practice when the calibrated estimators are used and the same weights are applied in estimating all the parameters needed.

We now formulate a proposition in which we present iterative equations for definition of the weights w_k.

Proposition 1. *The weights $w_k = w_k^{(i)}$, $k \in \mathbf{s}$, $i = 1, 3, 6, 7$, which satisfy* (4) *and minimize the distance function L_i, satisfy the equation $w_k^{(i)} = d_k g_k^{(i)}$. Here*

$$g_k^{(1)} = 1 + \lambda^{(1)} q_k e_k, \qquad g_k^{(3)} = \Big(\frac{1}{2}\lambda^{(3)} q_k e_k - 1\Big)^{-2},$$

$$g_k^{(6)} = 1 + \lambda^{(6)} d_k q_k e_k, \quad g_k^{(7)} = (\lambda^{(7)} d_k q_k e_k - 1)^{-2},$$

$$\lambda^{(1)} = \widehat{A}\Big(\sum_{k\in\mathbf{s}} d_k q_k e_k a_k b_k\Big)^{-1},$$

$$\widehat{A} = (N-1)Cov(a,b) + N\left(2 - \frac{\widehat{N}_w}{N}\right)\widehat{\mu}_{aw}\widehat{\mu}_{bw} - \sum_{k\in\mathbf{s}} d_k a_k b_k, \quad \widehat{N}_w = \sum_{k\in\mathbf{s}} w_k,$$

$$e_k = (a_k - \widehat{\mu}_{aw})(b_k - \widehat{\mu}_{bw}) - \left(1 - \frac{\widehat{N}_w}{N}\right)\left(\frac{\widehat{\mu}_{aw}}{a_k} + \frac{\widehat{\mu}_{bw}}{b_k}\right) a_k b_k;$$

$\lambda^{(3)}$ is a properly chosen root of the equation $\alpha_2\lambda^2 + \alpha_1\lambda + \alpha_0 = 0$ with

$$\alpha_0 = \widehat{A}, \quad \alpha_1 = -\sum_{k\in\mathbf{s}} q_k w_k a_k b_k e_k, \quad \alpha_2 = \frac{1}{4}\sum_{k\in\mathbf{s}} q_k^2 w_k a_k b_k e_k^2;$$

$\lambda^{(6)} = \widehat{A}\left(\sum_{k\in\mathbf{s}} d_k^2 q_k a_k b_k e_k\right)^{-1}$; $\lambda^{(7)}$ *is a properly chosen root of the equation $\beta_2\lambda^2 + \beta_1\lambda + \beta_0 = 0$ with*

$$\beta_0 = \widehat{A}, \quad \beta_1 = -2\sum_{k\in\mathbf{s}} d_k q_k w_k a_k b_k e_k \quad \beta_2 = \sum_{k\in\mathbf{s}} d_k^2 q_k^2 w_k a_k b_k e_k^2.$$

We introduce some additional notation: $c_k = (a_k - \mu_a)(b_k - \mu_b)$,

$$t_c = \sum_{k=1}^{N} c_k \quad \hat{t}_c = \sum_{k \in \mathbf{s}} d_k c_k.$$

Proposition 2. *The weights $w_k = w_k^{(i)}$, $k \in s$, $i = 1, 3, 6, 7$, which satisfy* (5) *and minimize the distance function L_i, satisfy the equation $w_k^{(i)} = d_k f_k^{(i)}$. Here*

$$f_k^{(1)} = 1 + (t_c - \hat{t}_c)\Big(\sum_{l \in \mathbf{s}} d_l q_l c_l^2\Big)^{-1} q_k c_k,$$

$$f_k^{(3)} = 4\bigg(2 - (N-1)Cov(a,b)\bigg(\sum_{l \in \mathbf{s}} \frac{q_l w_l^{3/2} c_l^2}{2(\sqrt{w_l} - \sqrt{d_l})}\bigg)^{-1} q_k c_k\bigg)^{-2},$$

$$f_k^{(6)} = 1 + (t_c - \hat{t}_c)\Big(\sum_{l \in \mathbf{s}} d_l^2 q_l c_l^2\Big)^{-1} d_k q_k c_k,$$

$$f_k^{(7)} = \bigg(1 - (N-1)Cov(a,b)\Big(\sum_{l \in \mathbf{s}} \frac{d_l q_l w_l^{3/2} c_l^2}{\sqrt{w_l} - \sqrt{d_l}}\Big)^{-1} d_k q_k c_k\bigg)^{-2}.$$

In the case of calibration of totals, the calibrated weights w_k of the estimator of covariance (3) are defined in the section 3.1.

Estimators of the population covariance using several systems of weights

Let us consider some other, more general estimators of the finite population covariance, which are constructed using several weighting systems. The new calibrated estimators of the covariance are of the following shape:

$$\widehat{Cov}_{mw}(y,z) = \frac{1}{N-1} \sum_{k \in \mathbf{s}} w_k^{[1]} \Big(y_k - \frac{1}{N} \sum_{l \in \mathbf{s}} w_l^{[2]} y_l\Big)\Big(z_k - \frac{1}{N} \sum_{l \in \mathbf{s}} w_l^{[3]} z_l\Big) \quad (7)$$

Several calibration equations may be used for definition of the calibrated weights $w_k^{[1]}$, $w_k^{[2]}$, $w_k^{[3]}$. Let us consider some of them.

Case 1. One can take a nonlinear equation

$$\widehat{Cov}_{mw}(a,b) = Cov(a,b). \quad (8)$$

Case 2. The systems of weights $w_k^{[1]}$, $w_k^{[2]}$, $w_k^{[3]}$ are defined by the following calibration equations:

$$\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k^{[1]}\big(a_k-\mu_a\big)\big(b_k-\mu_b\big) = Cov(a,b). \tag{9}$$

$$\sum_{k\in\mathbf{s}} w_k^{[2]} a_k = t_a, \quad \sum_{k\in\mathbf{s}} w_k^{[3]} b_k = t_b. \tag{10}$$

Case 3. The first system of weights $w_k^{[1]}$ is defined by the nonlinear calibration equation (4). Calibration equations (10) define the other two systems of the weights $w_k^{[2]}$ and $w_k^{[3]}$.

A reasonable choice of the distance function in the first three cases may be as follows:

$$L(w,d) = \alpha_1 \sum_{k\in\mathbf{s}} \frac{(w_k^{[1]}-d_k)^2}{d_k\, q_k} + \alpha_2 \sum_{k\in\mathbf{s}} \frac{(w_k^{[2]}-d_k)^2}{d_k\, q_k} + \alpha_3 \sum_{k\in\mathbf{s}} \frac{(w_k^{[3]}-d_k)^2}{d_k\, q_k}\,, \tag{11}$$

where $\alpha_1+\alpha_2+\alpha_3=1$, $0<\alpha_i<1$, $i=1,2,3$.

Case 4. We can consider the estimator of covariance which uses two systems of weights:

$$\widehat{Cov}_{mw}(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k^{[1]}\Big(y_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[2]} y_l\Big)\Big(z_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[2]} z_l\Big) \tag{12}$$

The first system of weights $w_k^{[1]}$ is defined by equation (9), whereas the second system $w_k^{[2]}$ satisfies the following equations

$$\sum_{k\in\mathbf{s}} w_k^{[2]} a_k = t_a, \quad \sum_{k\in\mathbf{s}} w_k^{[2]} b_k = t_b. \tag{13}$$

Case 5. We can use another combination of two systems of calibrated weights: the first one $w_k^{[1]}$ satisfies nonlinear calibration equation (4), where the system $w_k^{[2]}$ is defined by (13).

Case 6. The system of weights $w_k^{[1]}$ satisfies equation (9), whereas the system $w_k^{[2]}$ is equal to $w_k^{[3]}$, and it is obtained using nonlinear calibration equation (4).

The following distance function may be used for the last three cases:

$$L'(w,d) = \beta \sum_{k\in\mathbf{s}} \frac{(w_k^{[1]}-d_k)^2}{d_k\, q_k} + (1-\beta)\sum_{k\in\mathbf{s}} \frac{(w_k^{[2]}-d_k)^2}{d_k\, q_k}\,, \quad 0<\beta<1. \tag{14}$$

The first case is most complicated analytically, the expressions for the approximate iterative solutions of calibration equation (8) are complicated.

3.3. Estimation of variance of the calibrated estimators of population covariance

We consider in this section the problem of estimation of variance for the constructed calibrated estimators of population covariance. We have defined above three types of estimators having one weighting system. Due to the complexity of calibrated weights, the calculation of variance of the respective estimators is complicated. For example, in the case of a nonlinear calibration, calibrated weights are expressed by recurrent equations, and an explicit analytical expression for the variance is not available. The variance estimators in such a case are derived using a rough Taylor expansion, or some pseudo-linearization.

In some cases, for example, using calibration of known population totals and the distance function L_1, we obtain an explicit expression for the calibrated weights w_k. Consequently, we can derive, in such case, approximate variance for estimator of covariance.

Proposition 3. *The approximate variance of the estimator*

$$\widehat{Cov}(y,z)=\frac{1}{N-1}\sum_{k\in\mathbf{s}}w_k\Big(y_k-\frac{1}{N}\sum_{l\in\mathbf{s}}w_l y_l\Big)\Big(z_k-\frac{1}{N}\sum_{l\in\mathbf{s}}w_l z_l\Big),$$

where the weights w_k are defined using calibration of the known totals t_a, t_b and the distance function L_1 (1), is

$$AVar(\widehat{Cov}(y,z))=\frac{1}{(N-1)^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}e_k e_l,$$

$$\textit{where}\;\; e_k=y_k z_k-(\mu_z y_k-B_1 a_k)-(\mu_y z_k-B_2 b_k),$$

$$t_y=\sum_{k=1}^{N}y_k,\quad t_z=\sum_{k=1}^{N}z_k,$$

$$B_j=\boldsymbol{\tau}_j^{'}\mathbf{A}^{-1}\Big(-\mathbf{t}_{yz\mathbf{x}}+\frac{1}{N}(t_y\mathbf{t}_{z\mathbf{x}}+t_z\mathbf{t}_{y\mathbf{x}})\Big),\; j=1,2,$$

$$\boldsymbol{\tau}_1^{'}=(1,0),\quad \boldsymbol{\tau}_2^{'}=(0,1),\quad \mathbf{A}=\sum_{k=1}^{N}q_k\mathbf{x}_k\mathbf{x}_k^{'},$$

$$\mathbf{t}_{yz\mathbf{x}}=\sum_{k=1}^{N}q_k y_k z_k\mathbf{x}_k,\;\; \mathbf{t}_{y\mathbf{x}}=\sum_{k=1}^{N}q_k y_k\mathbf{x}_k,\;\; \mathbf{t}_{z\mathbf{x}}=\sum_{k=1}^{N}q_k z_k\mathbf{x}_k,\;\; \mathbf{x}_k^{'}=(a_k,b_k).$$

In the dissertation, we derive also the expression for approximate variance of estimator which is defined by the linear calibration equation (5) and distance function L_1.

3.4. Calibrated and model-calibrated estimators of the finite population covariance

In this section, two types of calibrated estimators of the finite population covariance are considered. The estimators of the first type are of the form (12). C. Wu and R.R. Sitter proposed the estimators of the second type. They are constructed by calibrating the sample design weights $d_{ij} = 1/\pi_{ij}$, $\pi_{ij} = P(i\&j \in \mathbf{s})$, and using the linear regression model. Below we recall these estimators, and at the end of this section, we introduce *adjusted model-calibrated* estimators of the population covariance, which are constructed using a slightly different linear regression model.

Consider a finite population $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$ of N elements. Suppose we have two study variables y and z defined on the population $\mathcal{U}$ and taking real nonnegative values. The values of the variables y and z are not known. Let a serve as an auxiliary variable for y and b for z.

Assuming that the relationship between y_i and $\mathbf{x}_i = (1, a_i, b_i)'$ (z_i and $\mathbf{x}_i = (1, a_i, b_i)'$) is described by the linear regression model ξ:

$$E_\xi(y_i) = \mathbf{x}_i^{'}\boldsymbol{\beta}, \quad E_\xi(z_i) = \mathbf{x}_i^{'}\boldsymbol{\gamma}, \tag{15}$$

C. Wu and R.R Sitter have obtained the linear regression model-assisted and calibrated estimator of the population covariance

$$\widehat{Cov}_{MC}(y, z) = \frac{1}{N(N-1)} \sum_{i\in\mathbf{s}} \sum_{j>i} w_{ij}(y_i - y_j)(z_i - z_j), \tag{16}$$

here the weights w_{ij} satisfy the following calibration equations

$$\frac{2}{N(N-1)} \sum_{i\in\mathbf{s}} \sum_{j>i} w_{ij} = 1, \tag{17}$$

$$\frac{1}{N(N-1)} \sum_{i\in\mathbf{s}} \sum_{j>i} w_{ij}(\hat{y}_i - \hat{y}_j)(\hat{z}_i - \hat{z}_j) = Cov(\hat{y}, \hat{z}) \tag{18}$$

and minimize distance function

$$L^\star = \sum_{i\in\mathbf{s}} \sum_{j>i} (w_{ij} - d_{ij})^2/d_{ij}. \tag{19}$$

The fitted values for y_i and z_i are denoted by $\hat{y}_i$ and $\hat{z}_i$.

For most populations considered, the accuracy of model-calibrated estimator (16) is lower than that of our calibrated estimators (12), which use two systems of weights. We modified the model-calibrated estimator of the population covariance (16) by introducing a new idea to consider the following regression model:

$$E_\xi(y_i) = \beta_0 + \beta_1 a_i, \quad E_\xi(z_i) = \gamma_0 + \gamma_1 b_i. \tag{20}$$

Using this model, calibration equations (17), (18), and distance function (19), we have derived a new estimator of the same form (16). We call it as *an adjusted model-calibrated* estimator.

The simulation results show that the adjusted model-calibrated estimator of covariance is more efficient compared to the estimator (16) provided the auxiliary variables are well correlated with the study variables.

General Conclusions

After solving the problems formulated in the section "The aim and problems of the dissertation", we have obtained the following results:

1. The expressions of the calibrated weights have been derived in the case of estimation of population totals. The results were obtained using seven different distance functions. The constructed calibrated estimators are more efficient compared to the straight estimators provided the auxiliary variables are well correlated with the study variables.

2. By virtue of different distance functions and calibration equations, three groups of calibrated estimators have been constructed for population covariance, using one, two, and three weighting systems, respectively. The results of mathematical simulation show that calibrated estimators of covariance are more efficient compared to the straight estimators provided the auxiliary variables are well correlated with the study variables. The increased number of weighting systems gives the reduced variance of estimators in many cases.

3. Several methods have been proposed for estimating the variance of constructed estimators of the population total and covariance (variance). The Taylor linearization technique played an important role here.

4. New Matlab functions have been created to compare by simulation the constructed calibrated estimators with the standard estimators of the respective parameters.

5. Using the linear regression model, an adjusted model-calibrated estimator of the finite population covariance has been constructed. In the case of

high correlated study and auxiliary variables, the mean square error of the adjusted estimator is smaller than the same characteristic of estimator, which was proposed by C. Wu and R.R. Sitter.

6. An adjusted geometric stratification method has been proposed for skewed populations. Simulation results show that the adjusted geometric stratification method outperforms all the methods in the case of highly skewed populations.

List of Published Works on the Topic of the Dissertation

In the reviewed scientific periodical publications

[A1] Plikusas, A.; Pumputis, D. Calibrated estimators of the population covariance. *Acta Applicandae Mathematicae*, 2007, Vol 97, No 1–3, p. 177–187. ISSN 0167-8019 (ISI Master Journal List).

[A2] Plikusas, A.; Pumputis, D. Calibrated estimators of totals under different distance measures. *Lietuvos matematikos rinkinys*, 2004, Vol 44, special issue, p. 572–576. ISSN 0132-2818.

[A3] Pumputis, D. Stratification of populations with skewed distribution. *Lietuvos matematikos rinkinys*, 2007, Vol 47, special issue, p. 369–374. ISSN 0132-2818.

[A4] Plikusas, A.; Pumputis, D. On the estimation of variance of calibrated estimators of the population covariance. *Statistics in Transition*, 2007, Vol 3, No 8, p. 475–486. ISSN 1234-7655 (IBSS).

In the other editions

[A5] Plikusas, A.; Pumputis, D. Estimation of the finite population covariance. In *Proceedings of the Eighth International Conference "Computer Data Analysis and Modeling: Complex Stochastic Data and Systems", held in Minsk, Belarus, on 11–15 September, 2007*. Minsk: BSU, 2007, p. 170–173. ISBN 978-985-476-508-2.

[A6] Pumputis, D. Calibrated estimators under different distance measures. In *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Vilnius on 17–21 June, 2005*. Vilnius: Statistics Lithuania, 2005, p. 137–141. ISBN 9955-588-87-X.

[A7] Plikusas, A.; Pumputis, D. Composite estimators for the sample and frame changing points. In *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Vilnius on 17–21 June, 2005*. Vilnius: Statistics Lithuania, 2005, p. 135–136. ISBN 9955-588-87-X.

[A8] Plikusas, A.; Pumputis, D. Calibrated estimators of finite population covariance. In *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Ventspils, Latvia, on 24–28 August, 2006*. Riga: Central Statistical Bureau of Latvia, 2006, p. 131–133.

[A9] Pumputis, D. On estimation of the variance of calibrated estimators of the population covariance. In *Proceedings of the Second Baltic-Nordic Conference on Survey Sampling, held in Kuusamo, Finland, on 2–7 June, 2007*. Helsinki: Statistics Finland, 2007, p. 59.

[A10] Krapavickaitė, D.; Plikusas, A.; Pumputis, D. Estimation of variance for the calibrated estimators of the finite population covariance. In *Proceedings of the 56th Session of the ISI, held in Lisboa, Portugal, on 22–29 August, 2007*. Lisboa, 2007 (available on CD).

[A11] Pumputis, D. Calibrated and model-calibrated estimators of the finite population covariance. In *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Kuressaare, Estonia, on 25–29 August, 2008*. Tallinn: Statistics Estonia, 2008, p. 138–143. ISBN 978-9985-74-451-2.

About the author

Dalius Pumputis was born in Zarasai, on 17 of September, 1980.

First degree in Mathematics, Faculty of Mathematics and Informatics, Vilnius Pedagogical University, 2002. Master of Science in Mathematics, Faculty of Mathematics and Informatics, Vilnius Pedagogical University, 2004. In 2004–2008 – PhD student of Institute of Mathematics and Informatics. In 2002–2004 was working as a technician, and in 2004–2008 – as an assistant at the Vilnius Pedagogical University, Faculty of Mathematics and Informatics. At present – lector at the Department of Algebra and Statistics, Faculty of Mathematics and Informatics, Vilnius Pedagogical University. Dalius Pumputis worked on probation twice at the Umeå University (Sweden) and once at the Stockholm University (Sweden).

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI, GAUTI NAUDOJANT PAPILDOMĄ INFORMACIJĄ

Tiriamoji problema

Šiuolaikinėje valstybinėje statistikoje ir kitose srityse yra įvairių papildomos informacijos šaltinių apie tiriamas populiacijas. Tai, pavyzdžiui, gyventojų, įmonių ar mokesčių inspekcijos registrų duomenys. Disertacijoje nagrinėjamos papildomos informacijos panaudojimo galimybės konstruojant baigtinės populiacijos sumos, dispersijos ir kovariacijos įvertinius, bei sluoksniuojant baigtines populiacijas.

Darbo aktualumas

Imčių metodai – jauna statistikos mokslo šaka, kuri sparčiai pradėjo vystytis po 1940 metų. Sovietų Sąjunga nedalyvavo plėtojant šią mokslo šaką, todėl Lietuvoje pirmieji rimti mokslo darbai, susiję su baigtinių populiacijų statistika, pasirodė tik po nepriklausomybės atgavimo. Šiandien imčių metodai, organizuojant apklausas, vertinant populiacijos parametrus, plačiai taikomi valstybinėje statistikoje, sociologiniuose, ekonominiuose ir kt. tyrimuose. Todėl svarbu plėtoti šią sritį, tobulinti esamus imčių metodus.

Daugelyje mokslo darbų pabrėžiama žinomos papildomos informacijos svarba konstruojant populiacijos parametrų įvertinius – jei papildomi kintamieji gerai koreliuoja su tyrimo kintamaisiais, tai pasitelkus tinkamus įvertinius, kurie naudoja papildomus kintamuosius, galima gauti tikslesnius įverčius. Ypač svarbus yra J.C. Deville ir C.E. Särndal straipsnis, kuriame pristatoma nauja baigtinės populiacijos sumos įvertinių, naudojančių papildomus kintamuosius, klasė – *kalibruotieji įvertiniai*. Šie įvertiniai vis plačiau taikomi valstybinėje statistikoje.

Be populiacijos sumos, egzistuoja daug kitų svarbių, bet sudėtingesnių parametrų: populiacijos dviejų sumų santykis, dispersija, kovariacija, kvantilis ir kt. Populiacijos dviejų sumų santykio įvertiniai gali būti pritaikyti darbo užmokesčio tyrimuose, kovariacijos įvertiniai – regresijos ir koreliacijos koeficientams bei kovariacinėms matricoms vertinti, o kvantilio įvertiniai – ekonomikos tyrimuose. Deja, nėra daug darbų, kuriuose būtų nagrinėjami šių parametrų įvertiniai, naudojantys žinomą papildomą informaciją. Taigi būtina praplėsti minėtų parametrų įvertinių klasę, pavyzdžiui, pritaikant J.C. Deville ir C.E. Särndal pasiūlytą svorių kalibravimo techniką.

Įverčių tikslumas bei jų dispersija priklauso ne tik nuo naudojamų įvertinių ar papildomos informacijos, bet ir nuo imties plano. Turint papildomos informacijos apie populiacijos struktūrą, dažnai efektyvus būna sluoksninis ėmimas, plačiai taikomas daugelyje tyrimų. Siekiant tikslesnių statistinio tyrimo rezultatų, reikia naudoti kuo efektyvesnius sluoksniavimo metodus, kurie gali būti gaunami tobulinant esamus ar kuriant naujus.

Darbo tikslas ir uždaviniai

Pagrindinis šio darbo tikslas – naudojantis papildoma informacija patobulinti tam tikrus baigtinės populiacijos sluoksniavimo ir populiacijos sumos bei kovariacijos vertinimo metodus. Siekiant tikslo buvo sprendžiami šie uždaviniai:

1. Naudojant skirtingas atstumo funkcijas bei kalibravimo lygtis sukonstruoti baigtinės populiacijos sumos, dispersijos ir kovariacijos kalibruotus įvertinius.

2. Sukonstruoti baigtinės populiacijos dispersijos ir kovariacijos kalibruotus įvertinius, naudojančius keletą svorių sistemų.
3. Sukonstruoti gautųjų kalibruotų įvertinių dispersijos įvertinius.
4. Remiantis matematiniu modeliavimu kalibruotus įvertinius palyginti su standartiniais atitinkamų parametrų įvertiniais.
5. Modeliuojant palyginti autoriaus (žr. [A1]) sukonstruotus kalibruotus kovariacijos įvertinius su C. Wu ir R.R. Sitter pasiūlytu tiesiniu regresiniu modeliu pagrįstu kalibruotu baigtinės populiacijos kovariacijos įvertiniu.
6. Modifikuoti tiesiniu regresiniu modeliu pagrįstą kalibruotą baigtinės populiacijos kovariacijos įvertinį kiekvienam tyrimo kintamajam naudojant atskirus papildomus kintamuosius.
7. Modifikuoti P. Gunning ir J.M. Horgan pasiūlytą geometrinį sluoksniavimo metodą, darant prielaidą, kad populiacijos skirstinys yra eksponentinis.

Tyrimų metodai

Darbe taikomi šie metodai: analizinis, tikimybinis ir eksperimentinis. Įrodant teiginius, taikyti Lagranžo daugiklių, Teiloro skleidimo eilute, atsitiktinių dydžių skaitinių charakteristikų skaičiavimo metodai ir matricų diferencijavimo taisyklės. Išvedant pataisytąjį tiesiniu regresiniu modeliu pagrįstą baigtinės populiacijos kovariacijos įvertinį, remtasi modeliais pagrįstų įvertinių teorija. Sprendžiant optimaliųjų sluoksnių ribų radimo uždavinį, nagrinėti šaknies iš f, geometrinis ir laipsnių sluoksniavimo metodai. Išvestas pataisytasis geometrinis populiacijų sluoksniavimo metodas yra pagrįstas populiacijos skirstiniu. Atliekant matematinį modeliavimą, naudotasi matematinių uždavinių sprendimo paketu Matlab.

Mokslinis darbo naujumas

Remiantis svorių kalibravimo idėja ir naudojant keletą skirtingų atstumo funkcijų, šioje disertacijoje sukonstruojami bei tarpusavyje empiriškai palyginami populiacijos sumos kalibruotieji įvertiniai. Naudodami tokias atstumo funkcijas, kurių išraiškoje yra, pavyzdžiui, kvadratinės šaknys, mes garantuojame, kad turėsime neneigiamus svorius, nes neigiami svoriai dažnai yra didesnės įvertinių dispersijos priežastis. Darbe taip pat pateikiamos sukonstruotų įvertinių apytikslės dispersijos.

Šios disertacijos kitas naujas rezultatas – baigtinės populiacijos kovariacijos (dispersijos) kalibruotieji įvertiniai, naudojantys vieną ar daugiau svorių sistemų. Svorių konstravimui pasitelktos naujos, specialiai šiam parametrui pritaikytos, kalibravimo lygtys. Disertacijoje taip pat sprendžiama sukonstruotų įvertinių dispersijų vertinimo problema. Tai – techniškai sudėtingas uždavinys, nes daugeliu atvejų kalibruotų svorių negalima užrašyti išreikštiniu pavidalu.

Pataisytasis tiesiniu regresiniu modeliu pagrįstas kovariacijos įvertinys – taip pat naujas rezultatas. Jis skiriasi nuo C. Wu ir R.R. Sitter pasiūlyto įvertinio tuo, kad jį konstruojant kiekvienam tyrimo kintamajam buvo naudotas atskiras papildomas kintamasis.

Šiame darbe taip pat sprendžiamas asimetrinių populiacijų sluoksniavimo uždavinys ir siūlomas *pataisytasis geometrinis sluoksniavimo metodas*.

Atliekant eksperimentus, buvo naudotasi paties autoriaus sukurtomis naujomis Matlab funkcijomis.

Ginamieji disertacijos teiginiai

1. Teiginys apie kalibruotųjų svorių išraišką, vertinant baigtinės populiacijos sumas.
2. Teiginiai apie kalibruotųjų svorių išraišką, vertinant baigtinės populiacijos dispersiją ir kovariaciją.
3. Kalibruotieji baigtinės populiacijos kovariacijos įvertiniai, naudojantys kelias svorių sistemas.
4. Teiginiai apie sukonstruotų kalibruotųjų populiacijos sumos ir kovariacijos (dispersijos) įvertinių apytikslių dispersijų skaičiavimą ir jų vertinimą.
5. Skirtingų kalibravimo lygčių įtakos vertinimo tikslumui analizė.
6. Pataisytojo geometrinio sluoksniavimo taisyklė asimetrinėse populiacijose.
7. Pataisytasis tiesiniu regresiniu modeliu pagrįstas kalibruotasis populiacijos kovariacijos įvertinys.

Darbo apimtis

Disertaciją sudaro įvadas, trys skyriai ir išvados. Papildomai disertacijoje yra pateikta vartotų žymėjimų ir santrumpų sąrašas, sąvokų žodynas, dalykinė rodyklė bei literatūros sąrašas. Darbo apimtis yra 134 puslapiai, kuriuose pateikta: 193 formulės, 6 histogramos ir 12 lentelių. Disertacijoje remtasi 116 literatūros šaltinių.

Bendrosios išvados

Išsprendus skyrelyje „Darbo tikslas ir uždaviniai“ suformuluotas problemas, gauti šie rezultatai:

1. Išvestos kalibruotųjų svorių išraiškos baigtinės populiacijos sumai vertinti. Rezultatai gauti naudojant septynias skirtingas atstumo funkcijas. Sukonstruoti įvertiniai yra kur kas tikslesni už standartinius įvertinius, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė.
2. Taikant skirtingas atstumo funkcijas ir kalibravimo lygtis, sukonstruoti trijų tipų baigtinės populiacijos kovariacijos kalibruotieji įvertiniai, naudojan-

tys atitinkamai vieną, dvi bei tris svorių sistemas. Matematinio modeliavimo rezultatai rodo, kad kalibruotieji kovariacijos įvertiniai yra tikslesni už standartinius įvertinius, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė. Svorių sistemų skaičiaus padidinimas iki 2 ar 3 daugeliu atvejų sumažina įvertinių dispersiją.

3. Pasiūlyti sukonstruotų kalibruotų populiacijos sumos, kovariacijos (dispersijos) įvertinių dispersijų vertinimo būdai. Svarbus vaidmuo čia tenka Teiloro ištiesinimo metodui.

4. Sukurtos naujos Matlab funkcijos, kurios atliekant šioje disertacijoje nagrinėtus eksperimentus buvo naudojamos kalibruotiesiems įvertiniams palyginti su standartiniais atitinkamų parametrų įvertiniais.

5. Sukonstruotas pataisytasis tiesiniu regresiniu modeliu pagrįstas kalibruotasis baigtinės populiacijos kovariacijos įvertinys. Modeliuojant pastebėta, kad šio įvertinio vidutinė kvadratinė paklaida yra mažesnė už C. Wu ir R.R. Sitter pasiūlyto įvertinio vidutinę kvadratinę paklaidą, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė.

6. Pasiūlytas pataisytasis geometrinis asimetrinių populiacijų sluoksniavimo metodas. Matematinio modeliavimo rezultatai rodo, kad šis metodas ypač asimetrinių populiacijų (kai $ac > 10$) atveju yra geriausias.

Trumpos žinios apie autorių

Dalius Pumputis gimė 1980 m. rugsėjo 17 d. Zarasuose.

2002 m. įgijo matematikos bakalauro laipsnį Vilniaus pedagoginio universiteto Matematikos ir informatikos fakultete. 2004 m. įgijo matematikos mokslo magistro laipsnį Vilniaus pedagoginio universiteto Matematikos ir informatikos fakultete. 2004–2008 m. – Matematikos ir informatikos instituto doktorantas. 2002–2004 m. dirbo vyr. laborantu, o 2004–2008 m. – asistentu Vilniaus pedagoginio universiteto Matematikos ir informatikos fakultete. Šiuo metu dirba lektoriumi Vilniaus pedagoginio universiteto Algebros ir statistikos katedroje. Dalius Pumputis stažavosi du kartus Umeå universitete (Švedija) ir vienąkart – Stokholmo universitete (Švedija).

Dalius Pumputis

STATISTICAL ESTIMATORS OF THE FINITE POPULATION PARAMETERS IN THE PRESENCE OF AUXILIARY INFORMATION

Summary of Doctoral Dissertation
Physical Sciences, Mathematics (01P)

Dalius Pumputis

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI, GAUTI NAUDOJANT PAPILDOMĄ INFORMACIJĄ

Daktaro disertacijos santrauka
Fiziniai mokslai, matematika (01P)

2008 12 29. 1,5 sp. l. Tiražas 100 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“, Saulėtekio al. 11, LT-10223 Vilnius
http://leidykla.vgtu.lt
Spausdino UAB „Baltijos kopija“,
Kareivių g. 13B, LT-09109 Vilnius
http://www.kopija.lt