VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Viktoras CHADYŠAS

STATISTICAL ESTIMATORS OF THE FINITE POPULATION PARAMETERS IN THE CASE OF SAMPLE ROTATION

SUMMARY OF DOCTORAL DISSERTATION
PHYSICAL SCIENCES,
MATHEMATICS (01P)

VGTU LEIDYKLA
Vilnius TECHNIKA 2009

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2005–2009.
Scientific Supervisor
Prof Dr Habil Leonas SAULIS (Vilnius Gediminas Technical University, Physical Sciences, Mathematics – 01P).
Consultant
Assoc Prof Dr Danutė KRAPAVICKAITĖ (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 Jonas Kazys SUNKLODAS (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P).
Members:
Prof Dr Habil Algimantas Jonas BIKELIS (Vytautas Magnus University, Physical Sciences, Mathematics – 01P),
Prof Dr Habil Romanas JANUŠKEVIČIUS (Vilnius Pedagogical University, Physical Sciences, Mathematics – 01P),
Assoc Prof Dr Mečislavas MEILŪNAS (Vilnius Gediminas Technical University, Physical Sciences, Mathematics – 01P),
Prof Dr Valentinas PODVEZKO (Vilnius Gediminas Technical University, Social Sciences, Economics – 04S).
Opponents:
Prof Dr Habil Mindaugas BLOZNELIS (Vilnius University, Physical Sciences, Mathematics – 01P),
Assoc Prof Dr Aleksandras PLIKUSAS (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 Senate Hall of Vilnius Gediminas Technical University at 2 p. m. on 28 January 2010.
Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel. +370 5 274 49 52, +370 5 274 49 56; fax +370 5 270 01 12;
e-mail: doktor@vgtu.lt
The summary of the doctoral dissertation was distributed on 24 December 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

Viktoras CHADYŠAS

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI ESANT IMTIES ROTACIJAI

DAKTARO DISERTACIJOS SANTRAUKA
FIZINIAI MOKSLAI,
MATEMATIKA (01P)

Vilnius LEIDYKLA TECHNIKA 2009

Disertacija rengta 2005–2009 metais Vilniaus Gedimino technikos universitete.
Mokslinis vadovas
prof. habil. dr. Leonas SAULIS (Vilniaus Gedimino technikos universitetas, fiziniai mokslai, matematika – 01P).
Konsultantas
doc. dr. Danutė KRAPAVICKAITĖ (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).
Disertacija ginama Vilniaus Gedimino technikos universiteto Matematikos mokslo krypties taryboje:
Pirmininkas
prof. habil. dr. Jonas Kazys SUNKLODAS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).
Nariai:
prof. habil. dr. Algimantas Jonas BIKELIS (Vytauto Didžiojo universitetas, fiziniai mokslai, matematika – 01P),
prof. habil. dr. Romanas JANUŠKEVIČIUS (Vilniaus pedagoginis universitetas, fiziniai mokslai, matematika – 01P),
doc. dr. Mečislavas MEILŪNAS (Vilniaus Gedimino technikos universitetas, fiziniai mokslai, matematika – 01P),
prof. dr. Valentinas PODVEZKO (Vilniaus Gedimino technikos universitetas, socialiniai mokslai, ekonomika – 04S).
Oponentai:
prof. habil. dr. Mindaugas BLOZNELIS (Vilniaus universitetas, fiziniai mokslai, matematika – 01P),
doc. dr. Aleksandras PLIKUSAS (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).

Disertacija bus ginama viešame Matematikos mokslo krypties tarybos posėdyje 2010 m. sausio 28 d. 14 val. Vilniaus Gedimino technikos universiteto senato posėdžių salėje.
Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (8 5) 274 49 52, (8 5) 274 49 56; faksas (8 5) 270 01 12;
el. paštas doktor@vgtu.lt
Disertacijos santrauka išsiuntinėta 2009 m. gruodžio mėn. 24 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“ 1703-M mokslo literatūros knyga.

Introduction

Scientific problem

Repeated sampling of population is a sampling procedure sometimes used in the official statistics. The construction of the estimators of the finite population parameters, such as finite population total, distribution function, quantile in the case of sample rotation is analyzed in the dissertation.

Topicality of the work

Survey sampling is a quite young branch of statistics, that has been rapidly developed since 1934. Nowadays, the methods of survey sampling are widely applied in the official statistics, public opinion, market research and other surveys. Therefore it is important to develop further methods of survey sampling.

Sometimes in social surveys the same population is sampled repeatedly on several occasions and the same study variable is observed in each occasion, so the development of the parameter estimated over time may be followed over time. Then information from the previous surveys is available and the estimation procedure can be combined with multi-phase sampling. The previous-phase sample data can be used as auxiliary information. In sample survey, supplementary information is often used at the estimation stage. If auxiliary variable is well correlated with the study variable, then it is possible to obtain more accurate estimate of parameter.

Except for the total of the values of study variable defined in finite population, there are a lot of other important, but complicated parameters: population variance, distribution function, quantile and others. Unfortunately, the estimation of these parameters, using auxiliary variables, is not widely studied in the literature, and particularly in the case of sample rotation. So, it is important to consider estimators of parameters mentioned above.

Research object

The research object of the work is estimation of the finite population parameters using auxiliary information in the case of sample rotation.

The aim and tasks of the work

The aim of this work is to propose some estimators of the finite population total, distribution function, quantile using auxiliary information in the case of sample rotation.

To achieve the aim of the work these tasks have to be solved:

1. To construct composite estimators of the finite population total and variance of these estimators using sample rotation. To compare the constructed estimators to standard estimator by simulation study, based on the real population data.

2. To construct composite estimators of the finite population distribution function and variance of these estimators using two occasion sampling scheme. To obtain optimal estimators in the sense of minimal variance. To compare standard estimator of the distribution function with the proposed estimators by simulation based on the real population data.

3. To obtain estimators of the finite population quantile by deriving the proposed distribution function estimators. To propose some procedures for estimation of the confidence intervals for finite population quantile using resampling methods.

Applied methods

The theory of design-based estimators is used for definition of the estimators of finite population parameters. The Taylor linearization technique, the methods for calculating numerical characteristics of random variables are used to prove the propositions formulated by the dissertation author. The resampling methods are used for construction of confidence intervals for finite population quantile. The statistical software SAS is used to perform all the simulations, described in this work.

Scientific novelty

We construct estimators of the finite population total, using auxiliary information known from previous survey in the case of sample rotation in this work. Composite ratio estimators of the population total are proposed. An approximate variances of constructed estimators are derived in this work also.

Estimation of the distribution function under sampling on two occasions is investigated. Composite regression and ratio type estimators are considered, using values of the study variables obtained on the first occasion, as auxiliary information. The optimal estimators, in the sense of minimal variance, are also obtain. Approximate variances formulas of constructed estimators are derived in this work as well.

We consider construction of the direct population quantile estimator with the use of auxiliary information. We are estimating population quantile using inverse of the estimator proposed of distribution function.

Some resampling methods using procedures are proposed for estimation of the confidence intervals for the finite population quantile.

Practical value of the work results

We propose some estimators of the population parameters which can be used in the social surveys of official statistics.

Statements presented for defence

1. Expressions of the composite estimators of the finite population total in the case of sample rotation.
2. The propositions on the expressions of approximate variance for the constructed composite ratio estimators of population total and its estimators.
3. Composite regression and ratio estimators of the finite population distribution function using two occasion sampling scheme.
4. The propositions on the expressions of approximate variance for the constructed composite regression and ratio estimators of population distribution function and its estimators.
5. Estimation of the finite population quantile using two occasion sampling scheme.
6. Estimation of confidence interval for the finite population quantile using resampling methods.

Approval of the work results

Six scientific papers have been published on the subject of the dissertation: four of them are in the reviewed scientific journals and two in the other scientific works. The research results were reported at the twelve scientific conferences also.

The scope of the scientific work

The scientific work layout consists of introduction, three chapters, conclusions, references and lists of authors publications. The total scope of the dissertation is 96 pages, 176 mathematical expressions, 6 figures, 10 tables and 56 items of references. The dissertation is written in Lithuanian.

1. Various estimators of the finite population parameters

Consider a finite population $\mathcal{U}=\{1, 2, \dots, N\}$ of N elements. Let y be a study variable defined on the population $\mathcal{U}$ and taking real values $y_1, \dots, y_N$. Let us study population parameter $\theta=\theta(y_1, \dots, y_N)$ which depends on study variable y. We are interested in estimation of the population parameter θ when probability sample $\mathfrak{s}$ of size n is drawn from the population $\mathcal{U}$ according to a given sampling design such that the inclusion probabilities $\pi_i = \mathrm{P}(i \in \mathfrak{s})$ and $\pi_{ij} = \mathrm{P}(i, j \in \mathfrak{s})$ are strictly positive.

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 estimators constructed using sample rotation.

In most surveys, interest centers on the current estimate, particulary if the characteristics of the population are likely to change rapidly with time. For current estimates, equal precision is obtained keeping the same sample or by changing it on every occasion. Replacement of part of the sample on each occasion may be better. The sample rotation can be performed as follows:

- Suppose population $\mathcal{U}$ is assumed to retain its composition over t time periods. At the $(t-1)$-th period, a sample $\mathfrak{s}'$ of size n' is extracted from $\mathcal{U}$ according to a certain sampling design and information is collected from all the elements of $\mathfrak{s}'$.

- A subsample $\mathfrak{s}_m$ of size m is selected from $\mathfrak{s}'$ according to a certain sampling design as well.

- This sample, $\mathfrak{s}_m$ is considered at the current t-th time period, when the rest of the sample, namely $\mathfrak{s}_u$ of size $u = n - n^{'}$ elements, are selected from the remaining elements $\mathfrak{s}'^{c} = \mathcal{U} \setminus s^{'}$ in accordance with a certain sampling design.

- The sample of the t-th occasion consists both of elements from $\mathfrak{s}_m$ and the ones from $\mathfrak{s}_u$. The estimators are constructed upon data of all these elements.

This scheme is named two occasion sampling scheme using data of t-th and $(t-1)$-th time periods. The problem of sampling on two occasions with partial replacement of sampling units was first considered by Jessen (1942). He provided the optimum estimate of the population mean $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$ on the second occasion in the case of simple random sampling on each of the two occasions

$$\overline{y}_t^{reg} = \alpha \overline{y}_{m,t}^{reg} + (1-\alpha)\overline{y}_{u,t}, \tag{1}$$

with

$$\alpha = \frac{Var(\overline{y}_{u,t})}{Var(\overline{y}_{m,t}^{reg}) + Var(\overline{y}_{u,t})}, \tag{2}$$

by combining two estimators of the mean: a double sampling regression estimator from the matched portion of the sample when the auxiliary variable is the value of the study variable y on the first occasion – $(t-1)$-th time period:

$$\overline{y}_{m,t}^{reg} = \overline{y}_{m,t} + b(\overline{y}_{n',t-1} - \overline{y}_{m,t-1}), \tag{3}$$

where b is linear regression coefficient of y in t-th and $t-1$-th time period, and an unbiased estimator $\overline{y}_{u,t}$ of the population mean is obtained from the unmatched sample $\mathfrak{s}_u$ on the t-th time period – second occasion.

The principle of deriving estimator (1) consists of improving the estimator which only uses the current information with the help of the correlations between the values of the study variable observed twice and belonging to the observations from the common sample. Composite estimation is used here as a term to describe estimators for a current period that use information from previous periods.

Sen *et al* (1975) has showed that the composite estimator of the population mean that uses a double sampling ratio estimator for the matched part of the sample out perform the unbiased estimator of the population mean when the auxiliary variable from the previous occasion is positively related to the study variable on current occasion, and their correlation coefficient is high enough.

Some problems of estimator construction for sampling on two occasions have been discussed in some details by Patterson (1950), Kulldorff (1963), Cochran (1977) and others. In all the studies, the parameter estimated is a mean. The problem of estimating more complex population parameters like quantile is quite different. Only recently this problem has been discussed (Martínez *et al* 2005; Rueda *et al* 2007).

2. Estimation of the finite population total in the case of sample rotation

In this chapter the estimation of the population total is considered in the case of sample rotation.

Labour force survey in official statistics of Lithuania is conducted quarterly, all members of a household are being interviewed in two subsequent quarters, allowed to rest for the next two, and returned to the sample for another two. For each quarter of the year one fourth of the previous quarter sample is changed by the new one. Sampling rotation scheme is shown in Fig 1. Despite information from the previous surveys is available, it is not used at the estimation stage at Statistics Lithuania yet.

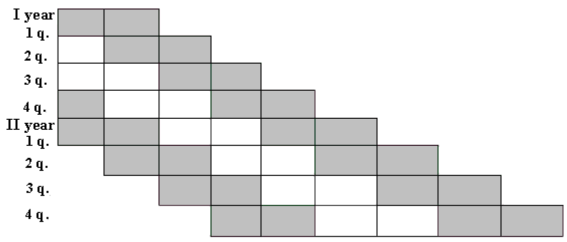

Fig. 1. Sample rotation sceme in labour force survey

Consider a finite population $\mathcal{U}$={$1, 2, \ldots, N$}, which is assumed to retain its composition over two–time periods. Let $y : y_1, \ldots, y_N$ be a study variable defined on the population $\mathcal{U}$. We are interested in the estimation of the population total of study variable y

$$t_y = \sum_{i \in \mathcal{U}} y_i.$$

2.1. Estimation of the total using scheme of two occasions

We are using two occasions sampling scheme for estimation population total t_y. Let us denote the study variable on the first occasion by x, with the values x_i, and the same variable on the second occasion by y with the values y_i.

Using the first occasion sample $\mathfrak{s}^{'}$ and the matched sample $\mathfrak{s}_m$, we can form a ratio estimator of the population total

$$\hat{t}^r_{ym} = \frac{\hat{t}_{yn'}}{\hat{t}_{xn'}} t_{xn'}. \tag{4}$$

A second design-based estimator $\hat{t}_{yu}$ of the population total t_y can be obtained from the unmatched sample $\mathfrak{s}_u$

We consider the composite ratio estimator of the total t_y under two occasion sampling scheme of the form

$$\hat{t}^{r*}_y = \alpha \hat{t}^r_{ym} + (1-\alpha)\hat{t}_{yu}, \tag{5}$$

where α, $\alpha \in (0, 1)$.

The variance of the linear part of the Taylor series of the estimator is called its approximate variance. We formulate a proposition in which the approximate variance of the composite ratio estimator $\hat{t}^{r*}_y$ is presented.

Proposition 1. *An approximate variance of composite ratio estimator $\hat{t}^r_y$ of the population total t_y is expressed:*

$$AVar(\hat{t}^{r*}_y) = \alpha^2 AVar(\hat{t}^r_{ym}) + (1-\alpha)^2 Var(\hat{t}_{yu}) + 2\alpha(1-\alpha)Cov(\hat{t}^r_{ym}, \hat{t}_{yu}). \tag{6}$$

Variance $Var(\hat{t}^r_{ym})$ of the ratio estimator $\hat{t}^r_{ym}$ may be expressed by conditional and unconditional variances and expectations:

$$Var(\hat{t}^r_{ym}) = Var\{E(\hat{t}^r_{ym} \mid \mathfrak{s}')\} + E\{Var(\hat{t}^r_{ym} \mid \mathfrak{s}')\}. \tag{7}$$

$\hat{t}^r_{ym}$ is nonlinear estimator, the approximate variance $AVar(\hat{t}^r_{ym})$ of $Var(\hat{t}^r_{ym})$ is derived using a linear term of its Taylor expansion.

Differentiating right hand expression in (6) with respect to coefficient α and equating the derivative to zero, we can derive the optimal value α_{opt} of the coefficient α.

Proposition 2. *The optimal value α_{opt} is*

$$\alpha_{opt} = \frac{Var(\hat{t}_{yu}) - Cov(\hat{t}^r_{ym}, \hat{t}_{yu})}{AVar(\hat{t}^r_{ym}) + Var(\hat{t}_{yu}) - 2Cov(\hat{t}^r_{ym}, \hat{t}_{yu})}. \tag{8}$$

Observe that α_{opt} depends on covariance, although some authors ignore it. Replacing the coefficient α by the coefficient α_{opt} in the ratio estimator $\hat{t}^{r*}_y$ (5) we obtain an optimal composite ratio estimator $\hat{t}^{r*}_{y\,opt}$ of the population total t_y.

The expressions for proposed composite estimator of the total and its approximate variance in the case of simple random sampling on each of the two occasions are obtained.

The simulation results with real data show that the proposed composite ratio estimator of population total is more efficient than the unbiased Horvitz-Thompson estimator. The use of covariance in (8) provide more accurate estimator.

2.2. Estimation of the total using more complex sampling scheme

We consider estimator of total using more complex sampling scheme than scheme of two occasions in the previous section. The sample selection scheme is described as follows:

Step 1. The sample $s^{(1)}_1$ of size $n^{(1)}$ is drawn from the population $\mathcal{U}$. Then sample $s^{(2)}_1$ of size $n^{(2)}$ is drawn from sample $s^{(1)}_1$.

Step 2. From the set $s^{(1)}_2 = \mathcal{U} \setminus s^{(1)}_1$ sample $s^{(2)}_2$ of size $m^{(2)}$ is drawn. The sample $s^{(3)}_2$ of size $m^{(3)}$ is drawn from sample $s^{(2)}_2$.

Step 3. Sample $s^{(2)}_3$ of size $u^{(2)}$ is drawn from the set $s^{(1)}_3 = \mathcal{U} \setminus (s^{(1)}_1 \cup s^{(2)}_2)$.

The total sample $\mathfrak{s}$ consists of the union of three samples: $\mathfrak{s} = \mathfrak{s}^{(2)}_1 \cup \mathfrak{s}^{(3)}_2 \cup \mathfrak{s}^{(2)}_3$. We consider three separate estimators of the population total t_y.

Case 1. The sample $\mathfrak{s}^{(2)}_1$ is obtained by a two-phase sampling:

$$\mathcal{U} \longrightarrow \mathfrak{s}^{(1)}_1 \longrightarrow \mathfrak{s}^{(2)}_1.$$

The values x_i, $i \in \mathfrak{s}^{(1)}_1$ of the study variable y in the first phase can be used as auxiliary information. We can form ratio estimator $\hat{t}^r_{y1}$ of the population total t_y

$$\hat{t}^r_{1y} = \hat{t}^{(1)}_{1x} \frac{\hat{t}^{(2)}_{1y}}{\hat{t}^{(2)}_{1x}}.$$

Case 2. The sample $\mathfrak{s}^{(3)}_2$ is considered as a three phase sample:

$$\mathcal{U} \longrightarrow \mathfrak{s}^{(1)}_2 = \mathcal{U} \setminus \mathfrak{s}^{(1)}_1 \longrightarrow \mathfrak{s}^{(2)}_2 \longrightarrow \mathfrak{s}^{(3)}_2.$$

The values x_i, $i \in \mathfrak{s}_2^{(2)}$ of the study variable y in the second phase can be used as auxiliary information. We can form ratio estimator $\hat{t}^r_{2y}$ of the population total t_y as well

$$\hat{t}^r_{2y} = \hat{t}^{(2)}_{2x} \frac{\hat{t}^{(3)}_{2y}}{\hat{t}^{(3)}_{2x}}.$$

Case 3. The sample $\mathfrak{s}_3^{(2)}$ is considered as a two-phase sample:

$$\mathcal{U} \longrightarrow \mathfrak{s}_3^{(1)} = \mathcal{U} \setminus (\mathfrak{s}_1^{(1)} \cup \mathfrak{s}_2^{(2)}) \longrightarrow \mathfrak{s}_3^{(2)}.$$

Because we do not have auxiliary information for elements in the first phase sample $\mathfrak{s}_3^{(1)}$, the population total t_y is estimated unbiasedly by the estimator $\hat{t}_{3y}$ using sample $\mathfrak{s}_3^{(2)}$ data.

We consider the composite ratio estimator of the total t_y under proposed sampling scheme of the form

$$\hat{t}^{r**}_y = \alpha \hat{t}^r_{1y} + \beta \hat{t}^r_{2y} + \gamma \hat{t}_{3y}, \tag{9}$$

where α, β and γ are constants $(0 < \alpha < 1, 0 < \beta < 1, 0 < \gamma < 1)$, satisfying $\alpha + \beta + \gamma = 1$.

We formulate a proposition in which approximate variance of the composite ratio estimator $\hat{t}^{r**}_y$ is presented.

Proposition 3. *An approximate variance of the composite ratio estimator* (9) *of the population total* t_y *is expressed by*

$$\begin{aligned} Var(\hat{t}^{r**}_y) &= \alpha^2 AVar(\hat{t}^r_{1y}) + \beta^2 AVar(\hat{t}^r_{2y}) + \gamma^2 Var(\hat{t}_{3y}) \\ &\quad + 2\alpha\beta Cov(\hat{t}^r_{1y}, \hat{t}^r_{2y}) + 2\alpha\gamma Cov(\hat{t}^r_{1y}, \hat{t}_{3y}) \\ &\quad + 2\beta\gamma Cov(\hat{t}^r_{2y}, \hat{t}_{3y}). \end{aligned} \tag{10}$$

An approximate variance $AVar(\hat{t}^r_{1y})$ of the ratio estimator $\hat{t}^r_{1y}$ is defined similarly as $AVar(\hat{t}^r_{ym})$ of ratio estimator $\hat{t}^r_{ym}$ in (6). The variance $Var(\hat{t}^r_{2y})$ of the ratio estimator $\hat{t}^r_{2y}$ for three phase sampling design is obtained by

$$\begin{aligned} Var(\hat{t}^r_{2y}) &= Var\{E(\hat{t}^r_{2y} \mid \mathfrak{s}_2^{(1)})\} + E\{Var(\hat{t}^r_{2y} \mid \mathfrak{s}_2^{(1)})\} \\ &= Var\{E[E(\hat{t}^r_{2y} \mid \mathfrak{s}_2^{(2)}) \mid \mathfrak{s}_2^{(1)}]\} + E\{Var[E(\hat{t}^r_{2y} \mid \mathfrak{s}_2^{(2)}) \mid \mathfrak{s}_2^{(1)}]\} \\ &\quad + E\{E[Var(\hat{t}^r_{2y} \mid \mathfrak{s}_2^{(2)}) \mid \mathfrak{s}_2^{(1)}]\}. \end{aligned} \tag{11}$$

Because $\hat{t}^r_{2y}$ is nonlinear estimator, the approximate variance $AVar(\hat{t}^r_{2y})$ is derived using a linear term of its Taylor expansion.

In the dissertation, we give also the expression for proposed composite estimator of the total and its approximate variances in the case of simple random sampling on each of the occasions.

The accuracy of the proposed estimators obtained is studied by modeling with real data. The simulation results show that composite estimator $\hat{t}_y^{r**}$ (9) of the population total t_y constructed using auxiliary information known for all quarters outperforms the unbiased Horvitz-Thompson estimator of the total of both estimators: $\hat{t}_y^{r*}$ (5) and $\hat{t}_{y\,opt}^{r*}$ in which the known auxiliary information is used for two quarters.

3. Estimation of the more complex parameters in finite population in the case of sample rotation

3.1. Estimation of the finite distribution function

Estimation of distribution function is an important objective in survey practice, when we want to identify groups of the population elements for which values of study variable lie substantially below or above some particular value. In this section, we investigate estimator of the distribution function of a study variable using sampling design on two occasions.

Consider a finite population $\mathcal{U} = \{1, 2, \ldots, N\}$. Let y be the study variable, defined on the population $\mathcal{U}$ and taking values $\{y_1, \ldots, y_N\}$. The values of the variable y are not known. We are interested in the estimation of the finite population distribution function of the study variable y:

$$F_y(z) = \frac{\sharp A_z}{N},$$

where for any given number z, $-\infty < z < \infty$, the set $A_z = \{l \in \mathcal{U} : \; y_l \leq z\}$ is indicated, and $\sharp A_z$ denotes the number of elements in the set A_z. Such a function $F_y(z)$ may be of considerable interest when y is a measure of income and the population consists of individuals or households.

Let us define an indicator variable $h(z)$ with the values

$$h_i(z) = \begin{cases} 1, & \text{if } y_i \leq z, \\ 0, & \text{if } y_i > z, \quad -\infty < z < \infty, \end{cases}$$

$i = 1, 2, \ldots, N$, and its total $t_{h(z)} = \sum_{i=1}^{N} h_i(z)$. Then the distribution function of the study variable y is expressed as the mean of values variable $h(z)$ in the population:

$$F_y(z) = \frac{t_{h(z)}}{N} = \frac{1}{N}\sum_{i=1}^{N} h_i(z). \tag{12}$$

Using sampling on two occasions we construct estimator of the distribution function with values x_i from the first occasion sample as auxiliary information.

Let us define a new indicator variable $g(z)$ with the values

$$g_i(z) = \begin{cases} 1, & \text{if } x_i \leq z, \\ 0, & \text{if } x_i > z, \end{cases}$$

$i = 1, 2, ..., N$, and the total $t_{g(z)} = \sum_{i=1}^{N} g_i(z)$. Then the distribution function $F_x(z)$ is expressed as

$$F_x(z) = \frac{t_{g(z)}}{N} = \frac{1}{N}\sum_{i=1}^{N} g_i(z). \tag{13}$$

Using the same ideas for construction of the composite estimator of the population total as in the Section 2.1, we write composite regression and composite ratio estimators of the distribution function.

I. Composite regression estimator

Using the first occasion sample $s^{'}$ and the matched sample s_m, we form a regression type estimator of the distribution function

$$\widehat{F}_{ym}^{reg}(z) = \frac{1}{N}\hat{t}_{h(z)_m}^{reg} = \frac{1}{N}\hat{t}_{h(z)_m} + \frac{1}{N}b(\hat{t}_{g(z)_{n'}} - \hat{t}_{g(z)_m}) \tag{14}$$

with

$$b = \frac{s_{h(z)g(z)}}{s^2_{g(z)}} = \frac{\sum_{i=1}^{N}(h_i(z) - \mu_{h(z)})(g_i(z) - \mu_{g(z)})}{\sum_{i=1}^{N}(g_i(z) - \mu_{g(z)})^2}.$$

We consider the composite regression estimator of the distribution function $F_y(z)$ under two occasions sampling scheme of the form

$$\widehat{F}_{y}^{reg}(z) = \omega\frac{1}{N}\hat{t}_{h(z)_m}^{reg} + (1-\omega)\frac{1}{N}\hat{t}_{h(z)_u}, \tag{15}$$

where ω is a constant $(0 < \omega < 1)$, $\hat{t}_{h(z)_u}$ is design-based estimator of the total $t_{h(z)_u} = \sum_{i=1}^{N} h_i(z)$ obtained from the unmatched sample $\mathfrak{s}_u$.

II. Composite ratio estimator

A particular case within the regression type estimator is the ratio estimator. Estimators of distribution function of the regression and ratio differ in the choice of the coefficient b in (14).

Using the first occasion sample $s^{'}$ and the matched sample s_m, we form a ratio estimator of the distribution function

$$\widehat{F}^{r}_{ym}(z) = \frac{1}{N}\hat{t}^{r}_{h(z)_m} = \frac{1}{N}\hat{t}_{g(z)_{n}{'}}\widehat{R}(z), \tag{16}$$

which corresponds to the choice $b = \frac{\hat{t}_{h(z)_m}}{\hat{t}_{g(z)_m}} = \widehat{R}(z)$.

We consider the composite ratio estimator of the distribution function $F_y(z)$ under two occasions sampling scheme of the form

$$\widehat{F}^{r}_{y}(z) = \lambda\frac{1}{N}\hat{t}^{r}_{h(z)_m} + (1-\lambda)\frac{1}{N}\hat{t}_{h(z)_u}, \tag{17}$$

where λ is a constant $(0 < \lambda < 1)$.

An approximate variance of composite estimators $\widehat{F}^{reg}_{y}(z)$ (15) and $\widehat{F}^{r}_{y}(z)$ (17), also optimal constants ω_{opt} and λ_{opt} are obtained in a similar way as for composite estimator t^{r*}_{y} (5) of the total t_y.

We give the expressions for proposed composite estimators of the distribution function and their approximate variances in the case of simple random sampling on each of the two occasions.

The proposed estimators of distribution function are analyzed by simulation study of the real population using the two occasions sampling scheme, with simple random sampling on each of the occasions. Different matching fractions m/n have been selected for simulation. We have calculated estimates of the distribution function of the study variable y at the points

$$z_1 = K_{0.10}, \quad z_2 = K_{0.25}, \quad z_3 = K_{0.50}, \quad z_4 = K_{0.75}, \quad z_5 = K_{0.90},$$

where K_q is the q-level quantile of the study variable y in the population. The simulation results show that the proposed composite estimators using auxiliary information can be used for improving the accuracy of the estimates of distribution function. The efficiency of the estimators proposed depends on the matching fraction and on the level of quantiles for two occasions sampling.

3.2. Estimators of the finite population quantile

In this section estimators of the quantile by taking the inverse of proposed composite distribution function estimators using two occasions sampling scheme is considered.

If the function $F(z)$ is continuous and increasing, the quantile K_{yq} of level q, $q \in (0,1)$ can be found as

$$K_{yq} = F_y^{-1}(q). \tag{18}$$

Here F_y^{-1} is the inverse function of F_y.

Because the distribution function $F_y(z)$ (12) of the study variable in a finite population is a step function, an inverse function of a finite population distribution function and quantile can be defined in some special way. Let us arrange the population elements in increasing order of the variable y: $y_1 \le y_2 \le \ldots \le y_N$. Then the inverse of finite population distribution function and the q level quantile K_{yq}, $0 < q < 1$, can be defined as

$$K_{yq} = F^{-1}(q) = \begin{cases} \frac{1}{2}(y_{k-1} + y_k), & \text{if } F(y_k) = q; \\ y_{k-1}, & \text{if } F(y_{k-1}) < q < F(y_k). \end{cases} \tag{19}$$

for some $k \in \mathcal{U}$.

In this work procedure of quantile estimation is implemented: an estimator $\widehat{F}_y(z)$ of the distribution function $F_y(z)$ is obtained, and then the quantile estimated by defining the inverse of distribution function estimator, i.e.,

$$\widehat{K}_{yq} = \widehat{F}^{-1}(q) = \begin{cases} \frac{1}{2}(y_{k-1} + y_k), & \text{if } \widehat{F}(y_k) = q; \\ y_{k-1}, & \text{if } \widehat{F}(y_{k-1}) < q < \widehat{F}(y_k), \end{cases} \tag{20}$$

This type of quantile estimator is called a direct quantile estimator. So, then the q level quantile K_{yq} of study variable y, using two occasion sampling scheme can be estimated by taking an inverse of proposed distribution function estimators (15) and (17), as

$$\widehat{K}^r_{yq} = \widehat{F}^{r-1}_y(q) \quad \text{and} \quad \widehat{K}^{reg}_{yq} = \widehat{F}^{reg-1}_y(q),$$

where $\widehat{F}^{r-1}_y(q)$ and $\widehat{F}^{reg-1}_y(q)$ is obtained in the same way as $\widehat{F}^{-1}(q)$ in (20).

Under two occasion sampling scheme Rueda *et al* (2007) propose indirect quantile composite ratio and regression estimators. The composite ratio estimator under two occasion sampling scheme is obtained by them by a linear combination of two estimators:

$$\widehat{K}^{r*}_{yq} = \omega \widehat{K}^r_{m,yq} + (1-\omega)\widehat{K}_{u,yq}. \tag{21}$$

Here

$$\widehat{K}^r_{m,yq} = \frac{\widehat{K}_{m,yq}}{\widehat{K}_{m,xq}} \widehat{K}_{n',xq},$$

where $\widehat{K}_{m,yq} = \widehat{F}^{-1}_{m,y}(q)$, $\widehat{K}_{m,xq} = \widehat{F}^{-1}_{m,x}(q)$ and $\widehat{K}_{n',xq} = \widehat{F}^{-1}_{n',x}(q)$.

The composite regression estimator is obtained in a similar way, replacing ratio estimator $\widehat{K}^r_{m,yq}$ by regression estimator

$$\widehat{K}^{reg}_{m,yq} = \widehat{K}_{m,yq} + b(\widehat{K}_{n',xq} - \widehat{K}_{m,xq}),$$

where b is a suitably chosen coefficient such that the variance of the proposed estimator $\widehat{K}^{reg}_{m,yq}$ is a minimum.

Assuming sampling on two occasions, we use simple random sampling design in each of the occasions. A simulation study, based on the real population data, is performed and the direct ratio type estimator $\widehat{K}^r_{yq}$ of population quantile K_{yq} is compared with traditional direct estimator $\widehat{K}_{yq}$ and indirect ratio type estimator K^*_{yq} of population quantile. Results show that effectiveness of direct and indirect ratio quantile estimators depend on matching fraction.

3.3. Estimation of confidence intervals for finite population quantile

In this section we consider the problem of estimation of confidence interval for the finite population quantile. Some procedures that may be used to obtain estimates of confidence intervals for quantiles in a finite population based on resampling methods in the case of simple random sampling are compared.

Consider a finite population $\mathcal{U} = \{1, 2, \ldots, N\}$ 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 $\mathfrak{s}$ of size n is drawn from the population $\mathcal{U}$ with a given sampling design.

Woodruff (1952) proposed a simple method of getting a $1-\alpha$ level confidence interval under general sampling design, using only the estimated distribution function and its standard error. If $\widehat{K}_{yq}$, is normally distributed, the $1-\alpha$ level confidence interval estimator for the population quantile is suggested to estimate by him by

$$\left(\widehat{F}^{-1}\left(q - z_{\alpha/2}\sqrt{\widehat{Var}\left(\widehat{F}(\widehat{K}_{yq})\right)}\right), \widehat{F}^{-1}\left(q + z_{\alpha/2}\sqrt{\widehat{Var}\left(\widehat{F}(\widehat{K}_{yq})\right)}\right)\right). \quad (22)$$

The confidence level of such an interval may be inaccurate for small sample size, because the distribution of sample quantile may be not well-approximated by a normal distribution. Alternatively, confidence intervals can be estimated using resampling techniques. We construct estimates of the confidence intervals for quantiles using resampling procedures in the case of simple random sampling.

The traditional bootstrap method (TB)

The procedure is:

- Draw a simple random sample of size n from a population.

- From the sample drawn, draw a simple random subsample with replacement of size n (bootstrap sample). Let its quantile be $\widehat{K}_{yq}^{(1)}$. Repeat this process B times, obtain quantiles $\widehat{K}_{yq}^{(1)}$, $\widehat{K}_{yq}^{(2)}$,..., $\widehat{K}_{yq}^{(B)}$.
- Taking the quantiles $\widehat{K}_{yq}^{\text{TB}}(\alpha/2)$ and $\widehat{K}_{yq}^{\text{TB}}(1-\alpha/2)$ of this set of the levels $\alpha/2$ and $1-\alpha/2$, we get the estimate of the confidence interval of the level $1-\alpha$ for the quantile K_{yq}.

In the same way estimates of the confidence intervals for quantiles are constructed using others resampling methods: rescaling bootstrap (Rao, Wu 1988), jackknife.

A simulation study based on two different artificial populations has been made to compare the performance of these intervals. It appears that the quality of the normal distribution based method (22) and traditional bootstrap method based estimates of the confidence interval for quantiles are similar. The estimates of the confidence interval obtained by the rescaling bootstrap and jackknife method are too narrow.

General conclusions

The following results are obtained solving the tasks under consideration:

1. Some composite ratio estimators have been constructed for population total, using auxiliary information from previous survey. Modeling with real data has been performed. Real data of the Labour force survey of Statistics Lithuania has been used. The results of simulation study show that composite estimators are more efficient compared to the traditional estimator.
2. Composite regression and ratio estimators of a distribution function, as well as optimal estimators, in the sense of minimizing the variance, for a two occasion sampling scheme have been proposed. Simulation has been performed using data of Statistics Lithuania. The simulation results show that proposed composite estimators using auxiliary information may be used to improve the accuracy of the estimates of distribution function. The efficiency of the estimators proposed depends on the matching fraction and on the level of quantiles.
3. Direct ratio type estimator of the population quantile has been proposed under two occasion sampling scheme. Proposed estimator is compared with indirect ratio estimator and traditional direct estimator of population quantile. Effectiveness of direct and indirect ratio type estimator of population quantile in comparison to the traditional direct estimator depends on matching fraction. Efficiency of direct and indirect ratio quantile esti-

mators behave in a similar way. Because calculation is more intensive for direct estimator of population quantile we recommended to choose indirect estimator of quantile.

4. Some estimates of confidence intervals for the finite population quantile have been calculated using some procedures of resampling methods. The simulation results show that in the case of simple random sampling the quality of the normal distribution-based method and traditional bootstrap method based estimates of the confidence interval for quantiles are similar.

References

Cochran, W. G. 1977. *Sampling Techniques*. New York: John Wiley and Sons, Inc. 3-rd edition. 428 p.

Jessen, R. J. 1942. Statistical investigation of a sample survey for obtaining farm facts, in *Iowa Agricultural Experimental Stafion Research Bulletin*, 304.

Kulldorff, G. 1963. Some problems of optimum allocation for sampling on two occasions, *Rev. Inst. Int. Stat.* 31: 24–57.

Martínez, M. D.; Rueda, M.; Arcos, A.; Román, Y.; Gónzalez, S. 2002. Quantile estimation under successive sampling, *Computational Statistics* 20: 385–399.

Patterson, H. D. 1950. Sampling on successive occasions with partial replacement of units, *Journal of the Royal Statistical Society. Series B Methodological* 12: 241–255.

Rao, J. N. K.; Wu, C. F. J. 1988. Resampling inference with complex survey data, *Journal of American Statistical Association* 83(401): 231–241.

Rueda, M.; Munoz, J. F.; Gonzales, S.; Arcos, A. 2007. Estimating quantiles under sampling on two occasions with arbitrary sample designs, *Comput. Stat. Data An.* 51: 6956–6613.

Sen, A. R.; Sellers, R.; Smith, G. E. J. 1975. The use of a ratio estimate in successive sampling, *Biometrics* 31: 673–683.

Woodruff, R. S. 1952. Confidence intervals for medians and other positions measures, *Journal of the American Statistical Association* 47: 635–646.

List of scientific publications on the topic of dissertation

In the reviewed scientific periodical publications

Chadyšas, V. 2009. Estimation of a total for rotated sample design using auxiliary information, *Lietuvos matematikos rinkinys. LMD darbai* 50: 281–286. ISSN 0132-2818.

Chadyšas, V. 2009. Estimation of a Distribution Function under Sampling on two Occasions, *Nonlinear Analysis: Modelling and Control* 14(3): 315–331. ISSN 1392-5113 [Index Copernicus].

Chadyšas, V. 2008. Estimation of Confidence Intervals for Quantiles in a Finite Population, *Mathematical Modelling and Analysis* 13(2): 195–202. ISSN 1648-3510 [Thomson ISI Web of Science]

Chadyšas, V.; Krapavickaitė, D. 2008. Dirbančių gyventojų skaičiaus vertinimas esant imties rotacijai (Estimation of employed persons in the case of sample rotation), *Lietuvos matematikos rinkinys. LMD darbai* 48/49:288–293. ISSN 0132-2818.

In the other editions

Chadyšas, V. 2009. Estimation of total using auxiliary information, in *Proceedings of the Baltic-Nordic-Ukrainian Summer School on Survey Statistics, held in Kyiv, Ukraine*, 53–57. ISBN 966-8725-02-6.

Chadyšas, V. 2008. Estimators of quantile for rotated sample design, in *Proceedings of the Baltic-Nordic workshop on Survey sampling theory and methodology, held in Kuressaare, Estonia*, 63–68. ISBN 978-9985-74-451-2.

About the author

Viktoras Chadyšas was born in Rokiškis, on 12 of April, 1981.

First degree in Informatics, Faculty of Fundamental Science, Vilnius Gediminas Technical University, 2002. Master of Applied Statistics, Faculty of Fundamental Sciences, Vilnius Gediminas University, 2004. In 2005–2009 – PhD student of Vilnius Gediminas Technical University, Department of Mathematical Statistics. In 2005–2007 – as an assistant at the Department of Mathematical Statistics, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University. From 2008 –lector at the Department of Mathematical Statistics, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University. Viktoras Chadyšas in 2008 was on internship at Stockholm University in Sweden.

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI ESANT IMTIES ROTACIJAI

Tiriamoji problema

Oficialiojoje statistikoje atliekami įvairūs statistiniai tyrimai, siekiant kuo tiksliau įvertinti įvairius baigtinės populiacijos parametrus. Disertacijoje konstruojami baigtinės populiacijos tyrimo kintamojo reikšmių sumos, tyrimo kintamojo pasiskirstymo funkcijos ir jo skirstinio kvantilio įvertiniai esant imties rotacijai.

Darbo aktualumas

Imčių metodai – jauna statistikos mokslo šaka, kuri sparčiai vystėsi po 1934-ųjų metų. Šiuo metu visi didžiausi statistiniai tyrimai atliekami taikant imčių metodus. Todėl svarbu plėtoti šią sritį, tobulinti esamus imčių metodus.

Socialiniuose ir kituose statistiniuose tyrimuose kartais iš turimos populiacijos imtys renkamos pakartotinai ir tas pats tyrimo kintamasis stebimas keliais laiko momentais. Informacija apie tyrimo kintamojo reikšmes gali būti žinoma iš ankstesnių imčių tyrimų. Tuomet populiacijos parametrų vertinimui gali būti pritaikytas kelių fazių ėmimas. Ankstesnės fazės duomenys gali būti naudojami kaip papildoma informacija.

Be populiacijos sumos yra daug kitų svarbių, bet sudėtingesnių parametrų: populiacijos dispersija, pasiskirstymo funkcija, kvantilis ir kt. Deja, nėra daug darbų, kuriuose būtų nagrinėjami šių parametrų įvertiniai, naudojant papildomą informaciją ir esant imties rotacijai. Taigi būtina nagrinėti minėtų parametrų įvertinius.

Tyrimų objektas

Darbo tyrimo objektas yra baigtinėje populiacijoje apibrėžto tyrimo kintamojo parametrų vertinimas naudojant imties rotaciją.

Darbo tikslas ir uždaviniai

Darbo tikslas – pasiūlyti baigtinės populiacijos tyrimo kintamojo reikšmių sumos, tyrimo kintamojo pasiskirstymo funkcijos ir jo skirstinio kvantilio įvertinius esant imties rotacijai.

Siekiant numatyto tikslo buvo sprendžiami šie uždaviniai:

1. Sudaryti baigtinės populiacijos tyrimo kintamojo sumos įvertinius, bei jų dispersijų įvertinius, esant imties rotacijai. Remiantis matematiniu modeliavimu ir naudojant realius duomenis palyginti siūlomus įvertinius su imties planu pagrįstu įvertiniu.

2. Sudaryti baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertinius bei pasiskirstymo funkcijos įvertinių atitinkamus dispersijos įvertinius, naudojant imties rotaciją. Sudaryti optimalius įvertinius, mini-

mizuojančius pasiskirstymo funkcijų įvertinių dispersijas. Naudojant realius duomenis palyginti imties planu pagrįstą pasiskirstymo funkcijos įvertinį su autoriaus sukonstruotais pasiskirstymo funkcijos įvertiniais.

3. Vertinti baigtinės populiacijos tyrimo kintamojo kvantilį naudojant šiame darbe pasiūlytų pasiskirstymo funkcijų įvertinių atvirkštinių funkcijų įvertinius. Pasiūlyti kelis kvantilių pasikliautinojo intervalo įvertinius, taikant kartotinių imčių metodus. Modeliuojant duomenis palyginti jų tikslumą.

Tyrimų metodai

Sudarant baigtinės populiacijos tyrimo kintamojo sumos, pasiskirstymo funkcijos, kvantilio įvertinius remtasi imties planu pagrįstų įvertinių teorija. Įrodant disertacijos teiginius taikyti Teiloro ištiesinimo, atsitiktinių dydžių skaitinių charakteristikų skaičiavimo metodai. Sudarant baigtinės populiacijos tyrimo kintamojo kvantilio pasikliautinojo intervalo įvertinius remtasi kartotinių imčių metodais. Matematinis modeliavimas atliktas naudojant statistinių programų paketą SAS.

Darbo mokslinis naujumas

Naudojant imties rotaciją ir žinomą papildomą informaciją iš ankstesnių imties rinkimų sudaromi baigtinės populiacijos tyrimo kintamojo reikšmių sumos sudėtiniai santykiniai įvertiniai yra tikslesni, lyginant su imties planu pagrįstu baigtinės populiacijos tyrimo kintamo reikšmių sumos įvertiniu.

Kitas disertacijos rezultatas – sudėtingesnių baigtinės populiacijos tyrimo kintamojo parametrų: pasiskirstymo funkcijos ir kvantilio įvertiniai, esant imties rotacijai. Pasiūlyti sudėtinis regresinis ir sudėtinis santykinis baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertiniai. Surasti optimalūs įvertiniai, kurių dispersija būtų mažiausia. Pateikiamos šių įvertinių apytikslių dispersijų išraiškos. Įvertintas baigtinės populiacijos tyrimo kintamojo skirstinio kvantilis naudojant sukonstruotų tyrimo kintamojo pasiskirstymo funkcijos įvertinių atvirkštines funkcijas.

Šiame darbe pasiūlyti keli baigtinės populiacijos kvantilio pasikliautinojo intervalo įvertiniai paremti imčių perrinkimo procedūromis.

Darbo rezultatų praktinė reikšmė

Darbe pasiūlyti įvairių baigtinės populiacijos parametrų įvertiniai gali būti taikomi įvairiuose oficialiosios statistikos statistiniuose tyrimuose.

Ginamieji teiginiai

1. Sudėtiniai santykiniai baigtinės populiacijos tyrimo kintamojo reikšmių sumos įvertiniai esant imties rotacijai.

2. Teiginiai apie sukonstruotų sudėtinių santykinių baigtinės populiacijos tyrimo kintamojo reikšmių sumos įvertinių apytikslių dispersijų skaičiavimą

ir jų vertinimą.

3. Sudėtiniai baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertiniai esant imties rotacijai.
4. Teiginiai apie sukonstruotų sudėtinių baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertinių apytikslių dispersijų skaičiavimą ir jų vertinimą.
5. Baigtinės populiacijos tyrimo kintamojo pasiskirstymo kvantilio vertinimas esant imties rotacijai.
6. Baigtinės populiacijos tyrimo kintamojo skirstinio kvantilio pasikliautinojo intervalo vertinimas naudojant kartotinių imčių metodus.

Darbo rezultatų aprobavimas

Disertacijos tema paskelbtos 6 mokslinės publikacijos iš jų keturios recenzuojamuose mokslo leidiniuose. Disertacijos rezultatai aptarti 12-oje konferencijų.

Disertacijos struktūra

Disertaciją sudaro įvadas, trys pagrindiniai skyriai, išvados, naudotos literatūros sąrašas ir autoriaus publikacijų disertacijos tema sąrašas. Pirmasis skyrius skirtas analitinei mokslinės literatūros disertacijos tema apžvalgai. Antrajame skyriuje pateikiami autoriaus sukonstruoti baigtinės populiacijos tyrimo kintamojo sumos įvertinai. Trečiajame skyriuje pateikiami autoriaus sukonstruoti baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos, jo skirstinio kvantilio ir kvantilio pasikliautinojo intervalo įvertiniai.

Darbo apimtis yra 96 puslapiai, tekste panaudotos 176 numeruotos formulės, 6 paveikslai ir 10 lentelių. Rašant disertaciją buvo pasinaudota 56 literatūros šaltiniais.

Bendrosios išvados

1. Sudaryti sudėtiniai santykiniai baigtinės populiacijos sumos įvertiniai, naudojant iš ankstesnių tyrimų žinomą papildomą informaciją esant imties rotacijai. Modeliavimas atliktas su realiais gyventojų užimtumo tyrimo duomenimis. Atlikti eksperimentai parodė, kad sudėtiniai santykiniai įvertiniai yra kur kas tikslesni už standartinį imties planu pagrįstą baigtinės populiacijos sumos įvertinį. Buvo pastebėta, kad ieškant optimalių koeficientų būtina įtraukti kovariacijos narį, nuo to įvertinių tikslumas tik pagerėja.
2. Sudaryti sudėtinis regresinis ir santykinis baigtinės populiacijos pasiskirstymo funkcijos įvertiniai, o taip pat ir optimalūs įvertiniai su mažiausia dispersija, naudojant dviejų ėmimų schemą. Su realiais pajamų ir gyventojų

sąlygų tyrimo duomenimis atliktas modeliavimas. Sudėtiniai pasiskirstymo funkcijos įvertiniai naudojantys papildomą informaciją iš ankstesnių rinkimų yra tikslesni už standartinį pasiskirstymo funkcijos įvertinį. Tokių įvertinių efektyvumas priklauso nuo imties dydžių antrajame ėmime, o taip pat nuo kvantilio lygio q.

3. Tiesioginis santykinis ir netiesioginis santykinis baigtinės populiacijos tyrimo kintamojo skirstinio q lygio kvantilio įvertinys su tradiciniu imties planu pagrįstu kvantilio įvertiniu naudojant dviejų ėmimų schemą palygintas empiriškai. Abu santykiniai įvertiniai gali būti tikslesni už tradicinį įvertinį, tačiau tam įtakos turi ir parenkami imčių dydžiai antrajame ėmime. Kadangi netiesioginį santykinį q lygio kvantilio įvertinį skaičiuoti yra paprasčiau, pranašumą galėtume teikti jam.
4. Sudaryti baigines populiacijos kvantilio pasikliautinojo intervalo įverčiai naudojant imčių perrinkimo procedūras. Tradicinės savirankos ir normaliuoju skirstiniu pagrįstu metodais konstruotų pasikliautinųjų intervalų įverčiai yra panašūs. Visrakčio metodu paremta siūloma procedūra nėra tinkama kvantilio pasikliautinajam intervalui vertinti.

Apie autorių

Viktoras Chadyšas gimė 1981 m. balandžio 12 d. Rokiškyje.

2003 m. įgijo informatikos bakalauro laipsnį Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakultete. 2005 m. įgijo taikomosios statistikos magistro laipsnį Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakultete. 2005–2009 m. – Vilniaus Gedimino technikos universiteto Matematinės statistikos katedros doktorantas. 2005 m. Už darbą „2005 Viktoro Chadyšo publikacijos“ matematikos, fizikos ir chemijos mokslų skyriuje buvo įteikta Lietuvos mokslų akademijos premija. 2005–2007 m. dirbo asistentu, o nuo 2008 m. dirba lektoriumi Vilniaus Gedimino Technikos universiteto Fundamentinių mokslų fakultete. Viktoras Chadyšas 2008 m. stažavosi Švedijos Stokholmo universitete.

Viktoras CHADYŠAS

STATISTICAL ESTIMATORS
OF THE FINITE POPULATION PARAMETERS
IN THE CASE OF SAMPLE ROTATION

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

BAIGTINĖS POPULIACIJOS PARAMETRŲ
STATISTINIAI ĮVERTINIAI
ESANT IMTIES ROTACIJAI

Daktaro disertacijos santrauka
Fiziniai mokslai, matematika (01P)

2009 12 17. 1,5 sp. l. Tiražas 70 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*