

## Dirbančių gyventojų skaičiaus vertinimas esant imties rotacijai

Viktoras CHADYŠAS, Danutė KRAPAVICKAITĖ (VGTU)

el. paštas: viktorasch@fm.vgtu.lt, krapav@ktl.mii.lt

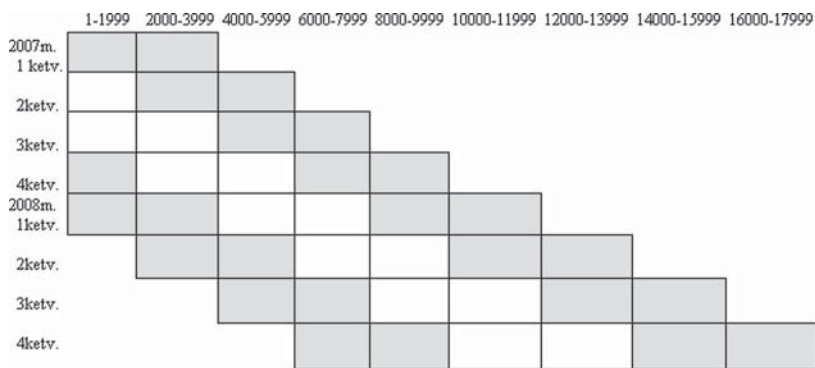
**Reziumė.** Straipsnelio tikslas – ištirti santykinio įvertinio taikymo efektyvumą gyventojų užimtumo tyrime naudojant paprastosios atsitiktinės imties planą su rotacija. Kiekvieną metų ketvirtį ketvirtadalis praėjusio ketvirčio imties keičiamas nauju, o trys ketvirtadaliai sudaromi iš anksčiau tirtų imčių. Šiek tiek supaprastinus imties planą konstruojamas sumos įvertinys, kuriame naudojama ankstesnio tyrimo papildoma informacija, ir šio įvertinio apytikslės dispersijos išraiška. Modeliuojant su realiais duomenimis tiriamas siūlomo įvertinio tikslumas.

**Raktiniai žodžiai:** gyventojų užimtumo tyrimas, dviejų fazių imties planas, santykinis įvertinys

### 1. Įvadas

Gyventojų užimtumo statistinis tyrimas – vienas svarbiausių daugelio šalių oficialiosios statistikos socialinių tyrimų. Jis atliekamas ir Statistikos departamente prie Lietuvos Respublikos Vyriausybės. Jame pagal pasirinktą imties planą iš gyventojų registro išrinkus 14 m. ir vyresnių gyventojų imtį, iš jų surenkami duomenys ir vertinamas dirbančių gyventojų, bedarbių ir neaktyvių gyventojų skaičius Lietuvoje ir jos dalyse. Gyventojų užimtumo statistinis tyrimas atliekamas kiekvieną metų ketvirtį.

Kadangi norima sekti užimtumo rodiklių kitimą laike, tai, siekiant sumažinti įverčių svyravimą dėl imties pakeitimo, pereinant nuo vieno ketvirčio prie kito keičiama



1 pav. Gyventojų užimtumo statistinio tyrimo imties rotacijos schema.

ne visa imtis, o tiksliai jos ketvirtadalis, t.y. daroma imties rotacija. Šiek tiek supaprastinus situaciją, galima pasakyti, kad kiekvienas į imtį išrinktas asmuo įtraukiamas į dviejų iš eilės atliekamų tyrimų imtis, po to į dviejų tyrimų imtį jis netraukiamas ir vėl įtraukiamas į dviejų iš eilės atliekamų statistinių tyrimų imtis. Imties rotacijos schema vaizduojama 1 pav. Horizontaliai žymimi imties elementų sąlyginiai kodai (numeriai), vertikaliai – statistinio tyrimo laikas.

Dėl imties rotacijos apie kiekvieną imties elementą, kuris įtrauktas į imtį jau nebe pirmą kartą, turimi ne tiksliai tiriama metų ketvirčio duomenys, bet ir anksčiau atlikto tyrimo duomenys. Kol kas skaičiuojant užimtumo parametrų įverčius Statistikos departamente turima ankstesnių tyrimų informacija apie imties elementus nebuvo naudojama. Remiantis bendrąja imčių iš baigtinės populiacijos teorija ([1]–[3]), santykiniai įvertiniai, kuriuose naudojama papildoma informacija iš ankstesnių tyrimų, galėtų pagerinti reikiamų įverčių tikslumą. Šio darbo tikslas – taikant santykinus įvertinius rotuojamai imčiai ir ankstesnio ketvirčio duomenis, modeliuojant su realiais duomenimis patikrinti šių įvertinių efektyvumą.

## 2. Populiacija ir imtis

Tarkime, turime baigtinę populiaciją  $U = \{1, \dots, N\}$ , kurios elementų tyrimo kintamojo reikšmės duotu momentu yra  $y_1, \dots, y_N$ . Tarkime, ankstesniame statistiniame tyrime iš populiacijos buvo išrinkta  $n'$  dydžio paprastoji atsitiktinė imtis  $\mathbf{i}'$  su tikimybe  $p(\mathbf{i}') = 1/C_N^{n'}$ . Jos elementų kintamojo  $y$  reikšmės ankstesniame tyrime žymėsime raide  $x$  su reikšmėmis  $x_k, k \in \mathbf{i}'$ .

Dabartiniam tyrimui renkamas paprastasis atsitiktinis  $m$  dydžio poimtis  $\mathbf{i}_m, \mathbf{i}_m \subset \mathbf{i}'$  iš ankstesnio statistinio tyrimo imties ir  $u$  dydžio poimtis  $\mathbf{i}_u \subset \mathbf{i}^c = U \setminus \mathbf{i}'$  iš populiacijos elementų, kurie nebuvo išrinkti į ankstesnį tyrimą. Tokiu būdu, nagrinėjama  $n = m + u$  dydžio imtis  $\mathbf{i} = \mathbf{i}_m \cup \mathbf{i}_u$ , gauta apjungus dvi dviejų fazių imtis su paprastosios atsitiktinės imties planais kiekvienoje fazėje.

Norėdami įvertinti dirbančių gyventojų skaičių Lietuvoje, laikykime tyrimo kintamąjį  $y$ , o taip pat ir  $x$ , dvireikšmiu:

$$y_k = \begin{cases} 1, & \text{jei gyventojas } k \text{ dirba,} \\ 0 & \text{priešingu atveju,} \end{cases} \quad x_k = \begin{cases} 1, & \text{jei gyventojas } k \text{ dirbo,} \\ 0 & \text{priešingu atveju,} \end{cases}$$

$k \in U$ . Tada dirbančių gyventojų skaičius lygus kintamojo  $y$  reikšmių populiacijos sumai:

$$t_y = \sum_{k \in U} y_k.$$

Vertinsime šią sumą, naudodami nurodytos imties kintamojo  $y$  reikšmes  $y_k, k \in \mathbf{i}$ , ir papildomo kintamojo  $x$  – ankstesnio tyrimo kintamojo – reikšmes  $x_k, k \in \mathbf{i}'$ .

## 3. Įvertinys ir jo dispersija

Pažymėkime pirmosios fazės imties  $\mathbf{i}'$  elementų kintamųjų  $y$  ir  $x$  sumas taip:

$$t_{yn'} = \sum_{k \in \mathbf{i}'} y_k, \quad t_{xn'} = \sum_{k \in \mathbf{i}'} x_k.$$

Suma  $t_{xn'}$  yra žinoma iš ankstesnių statistinio tyrimo rezultatų.

Iš antrosios fazės imties  $\mathbf{i}_m$  duomenų vertinkime šias sumas nepaslinktaisiais įvertiniais:

$$\hat{t}_{yn'} = \frac{n'}{m} \sum_{k \in \mathbf{i}_m} y_k, \quad \hat{t}_{xn'} = \frac{n'}{m} \sum_{k \in \mathbf{i}_m} x_k.$$

Tada galėsime įvertinti sumą  $t_{yn'}$  santykiniais įvertiniais

$$\hat{t}_{yn'}^{san} = \frac{\hat{t}_{yn'}}{\hat{t}_{xn'}} t_{xn'},$$

o sumą  $t_y$  – įvertiniais

$$\hat{t}_{ym}^{san} = \frac{N}{n'} \hat{t}_{yn'}^{san}. \quad (1)$$

Jei kintamieji  $y$  ir  $x$  pakankamai stipriai koreliuoti, tai, kaip rodo teorija ([1]–[3]), santykinis įvertinys  $\hat{t}_{yn'}^{san}$  gali turėti mažesnę dispersiją negu tiesioginis, imties planu pagrįstas, sumos  $t_{yn'}$  įvertinys  $\hat{t}_{yn'}$ . O dabar sudarykime imties planu pagrįstą populiacijos sumos  $t_y$  įvertinį dviejų fazių imties atveju, naudojant poimčio  $\mathbf{i}_u$  duomenis:

$$\hat{t}_{yu} = \frac{N}{N-n'} \left( \frac{N-n'}{u} \sum_{k \in \mathbf{i}_u} y_k \right) = \frac{N}{u} \sum_{k \in \mathbf{i}_u} y_k. \quad (2)$$

Iš (1) ir (2) įvertinių sudaromas sudėtinis populiacijos sumos  $t_y$  įvertinys:

$$\hat{t}_y = \alpha \hat{t}_{ym}^{san} + (1-\alpha) \hat{t}_{yu}, \quad (3)$$

čia  $\alpha$ ,  $\alpha \in (0, 1)$  – laisvai pasirenkama konstanta. Sudėtiniame įvertinyje naudojami visos imties  $\mathbf{i}$  duomenys.

**TEIGINYS.** Populiacijos sumos  $t_y$  įvertinio  $\hat{t}_y$  (3) apytikslė dispersija išreiškiama

$$AD\hat{t}_y = \alpha^2 AD\hat{t}_{ym}^{san} + (1-\alpha)^2 D\hat{t}_{yu} + 2\alpha(1-\alpha) Cov(\hat{t}_{ym}^{san}, \hat{t}_{yu}). \quad (4)$$

Čia naudojami žymenys:

1) Dviejų fazių imties atveju santykinio įvertinio apytikslė dispersija

$$AD\hat{t}_{ym}^{san} = N^2 \left(1 - \frac{n'}{N}\right) \frac{s_y^2}{n'} + N^2 \left(1 - \frac{m}{n'}\right) \frac{s_r^2}{m},$$

$$s_y^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - \mu_y)^2, \quad \mu_y = t_y/N,$$

$$s_r^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - r x_k)^2, \quad r = \sum_{i=1}^N y_i / \sum_{i=1}^N x_i.$$

2) Nepaslinktojo sumos įvertinio dispersija dviejų fazių imties plano atveju

$$D\hat{t}_{yu} = N^2 \left(1 - \frac{u}{N}\right) \frac{s_y^2}{u}.$$

3) Abiejų sumos įvertinių kovariacija  $Cov(\hat{t}_{ym}^{san}, \hat{t}_{yu}) \cong -Ns_y^2$ .

Pažymėkime  $D_1 = AD\hat{t}_{ym}^{san}$ ,  $D_2 = D\hat{t}_{yu}$ ,  $C = Cov(\hat{t}_{ym}^{san}, \hat{t}_{yu})$ . Minimizuojant apytikslę dispersiją  $AD\hat{t}_y$  ir laikant imties dydžius  $u, m$  pastoviais, randama optimali koeficiento  $\alpha$  reikšmė

$$\alpha_{opt} = \frac{D_2 - C}{D_1 + D_2 - 2C}.$$

Tada su koeficientu  $\alpha_{opt}$  gauto (3) optimalaus sumos įvertinio  $\hat{t}_{y opt}$  apytikslę dispersija turi pavidalą

$$AD(\hat{t}_{y opt}) = \frac{D_1 D_2 - C^2}{D_1 + D_2 - 2C}.$$

Vertinant apytikslę dispersiją (4), naryje  $AD\hat{t}_{ym}^{san}$  kintamojo  $y$  populiacijos dispersijos  $s_y^2$  ir  $s_r^2$  keičiamos į

$$\hat{s}_{ym}^2 = \frac{1}{m-1} \sum_{i \in \mathbf{i}_m} (y_k - \bar{y}_m)^2, \quad \bar{y}_m = \frac{1}{m} \sum_{k \in \mathbf{i}_m} y_k,$$

$$\hat{s}_r^2 = \frac{1}{m-1} \sum_{i \in \mathbf{i}_m} (y_k - \hat{r}x_k)^2, \quad \hat{r} = \frac{\sum_{k \in \mathbf{i}_m} y_k}{\sum_{k \in \mathbf{i}_m} x_k}.$$

Naryje  $D\hat{t}_{yu}$  kintamojo  $y$  populiacijos dispersija  $s_y^2$  keičiama į

$$\hat{s}_{yu}^2 = \frac{1}{u-1} \sum_{k \in \mathbf{i}_u} (y_k - \bar{y}_u)^2, \quad \bar{y}_u = \frac{1}{u} \sum_{k \in \mathbf{i}_u} y_k,$$

naryje  $Cov(\hat{t}_{ym}^{san}, \hat{t}_{yu})$  kintamojo  $y$  populiacijos dispersija  $s_y^2$  keičiama į

$$\hat{s}_y^2 = \frac{1}{n-1} \sum_{k \in \mathbf{i}} (y_k - \bar{y})^2, \quad \bar{y} = \frac{1}{n} \sum_{k \in \mathbf{i}} y_k.$$

#### 4. Modeliavimo rezultatai

Modeliavimui naudojami Statistikos departamento 2005 m. 1-ojo ir 2-ojo ketvirčių gyventojų užimtumo statistinio tyrimo duomenys. Abiejų ketvirčių imtims priklausantys elementai laikomi tyrimo populiacija, kurios dydis  $N = 4485$ .

1 lentelė. Modeliavimo populiacijos dviejų gretimų metų ketvirčių duomenys

Ankstesnis tyrimas	Dabartinis tyrimas	Nedirbantieji	Dirbantieji	Iš viso
Nedirbantieji		2 452	61	2 513
Dirbantieji		38	1 934	1 972
Iš viso		2 490	1 995	4 485

2 lentelė. Dirbančių, bedarbių ir neaktyvių gyventojų skaičiaus įverčiai ir jų variacijos koeficientų įverčiai

	$t_y$	$\hat{t}_{y\ PAI}$	$\tilde{c}\tilde{v}(\hat{t}_{y\ PAI})$	$\hat{t}_y$	$\tilde{c}\tilde{v}(\hat{t}_y)$	$\hat{t}_{y\ opt}$	$\tilde{c}\tilde{v}(\hat{t}_{y\ opt})$
Dirbantieji gyventojai	1 995	1 980	0.054	1 998	0.046	2 002	0.044
Bedarbiai	183	161	0.262	186	0.214	195	0.200
Neaktyvūs gyventojai	2 307	2 323	0.046	2 303	0.040	2 299	0.038

Iš 1 lentelės matome, kad praėjus trims mėnesiams iš 2 513 nedirbusių asmenų tik 61 pradėjo dirbti, ir iš 1 972 dirbusių tik 38 nustojo dirbti. Tai reiškia, kad tyrimo kintamojo reikšmių pasikeitimai laike yra nedideli, todėl galima tikėtis efektyvaus dirbančių gyventojų skaičiaus įverčio  $\hat{t}_y$ . Modeliavimui naudojami imties dydžiai  $n' = n = 400$ ,  $m = u = 200$ . Pakartojus imties išrinkimą ir įverčio  $\hat{t}_y$  skaičiavimą  $M = 100$  kartų, apskaičiuojamas empirinis įverčių vidurkis ir jų dispersijų vidurkis:

$$\bar{\hat{t}}_y = \frac{1}{M} \sum_{i=1}^M \hat{t}_y, \quad \tilde{D}(\hat{t}_y) = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{D}}(\hat{t}_y).$$

Skaičiuojamas empirinis įvertinio  $\hat{t}_y$  variacijos koeficientas:

$$\tilde{c}\tilde{v}(\hat{t}_y) = \frac{\sqrt{\tilde{D}(\hat{t}_y)}}{\hat{t}_y} \cdot 100(\%).$$

Palyginimui skaičiuojamas nepaslinktasis įvertinys  $\hat{t}_{y\ PAI} = N \sum_{k \in I} y_k / n$  ir vertinama jo dispersija  $\mathbf{D}\hat{t}_{y\ PAI} = N^2(1 - n/N)s_y^2/n$ , kurie būtų gaunami, jei  $n$  dydžio imtis būtų paprastoji atsitiktinė. Skaičiavimo rezultatai pateikiami 2 lentelėje. Joje  $\hat{t}_y$  žymimas įvertinys su  $\alpha = 1/2$ , o  $\hat{t}_{y\ opt}$  – įvertinys su įvertinta  $\alpha_{opt}$  reikšme.

## 5. Išvados

Iš modeliavimo rezultatų matome, kad dirbančiųjų ir neaktyvių gyventojų skaičiaus įverčių variacijos koeficientai yra žymiai mažesni negu bedarbių skaičiaus įverčio.

Taip yra todėl, kad bedarbiai sudaro žymiai mažesnę ir populiacijos, ir imties dalį, negu dirbantieji arba neaktyvūs gyventojai. Iš modeliavimo rezultatų taip pat matome, kad sudėtinio įvertinio  $\hat{t}_y$ , gaunamo naudojant papildomą informaciją dviejų fazių imties plano atveju, variacijos koeficiento įvertis sumažėjo, lyginant su nepaslinktuojų įvertiniu  $\hat{t}_{y\text{ PAI}}$ .

Optimalaus koeficiento  $\alpha$  naudojimas sudėtiniame įvertinyje (3) leidžia šio įvertinio variacijos koeficientą dar labiau sumažinti, tačiau tai gaunama padidėjusio poslinkio sąskaita.

Realiam gyventojų užimtumo statistinio tyrimo imties planui tikslinga naudoti papildomą informaciją santykiniam įvertiniui visiems trims metų ketvirčiams, kuriems turime ankstesnių tyrimų duomenų. Nuo to įvertinių tikslumas turėtų dar labiau pagerėti.

## 6. Padėka

Autoriai dėkoja VGTU studentui Andrejui Simanovičiui už šaunų indėlį atliekant šį darbą ir Statistikos departamentui už galimybę naudoti statistinio tyrimo duomenis.

## Literatūra

1. W.G. Cochran, *Sampling Techniques*, John Wiley & Sons, New York (1977).
2. D. Krapavickaitė, A. Plikusas, *Imčių teorijos pagrindai*, Technika, Vilnius (2005).
3. C.-E. Särndal, B. Swensson, J. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, New York (1992).

## SUMMARY

### *V. Chadyšas, D. Krapavickaitė. Estimation of employed persons in the case of sample rotation*

The aim of the paper – to investigate effectiveness of the ratio estimator in the labour force survey in the case of the simple random sampling with rotation. For each quarter of the year one fourth of the previous quarter sample is changed with the new one, and three fourth's are remaining the same. After simplification of the sampling design the estimator using auxiliary information and its approximate variance is constructed. The accuracy of the estimator obtained is studied by modelling with the real data.

*Keywords:* labour force survey, two phase sampling design, ratio estimator.