# LIETUVOS APSKRIČIŲ IR SAVIVALDYBIŲ BEDARBIŲ IR UŽIMTŲ GYVENTOJŲ SKAIČIAUS VERTINIMAS TAIKANT MODELIUS

**Linas Naujanis[1], Danutė Krapavickaitė[2]**

[1]Lietuvos statistikos departamentas
[2]Vilniaus Gedimino technikos universitetas
Addresas: [1]Gedimino pr. 29, LT-01500 Vilnius, Lietuva
[2] Saulėtekio al. 11, LT-10223 Vilnius, Lietuva
El. paštas: [1]linas.naujanis@stat.gov.lt, [2]danute.krapavickaite@vgtu.lt

Gauta: 2015 m. rugsėjis Pataisyta: 2015 m. spalis Paskelbta: 2015 m. lapkritis

**Santrauka.** Darbe nagrinėjami baigtinės populiacijos parametrų vertinimo srityse uždaviniai. Parametrams įvertinti taikomi keturi metodai: imties planu pagrįstas nepaslinktasis įvertinys, logistinės regresijos ir sričių lygio daugialypės regresijos modeliais pagrįsti įvertiniai, taip pat Džeimso ir Štaino įvertinys. Imties planu pagrįstas įvertinys yra nepaslinktasis, bet jo dispersija paprastai yra didelė. Modeliais pagrįsti įvertiniai nėra nepaslinktieji, bet turi mažesnes dispersijas. Tam, kad būtų sumažinta ir dispersija, ir sumažėtų poslinkis, taikomas Džeimso ir Štaino įvertinys. Baigtinės populiacijos parametrų srityse vertinimui naudojami Lietuvos statistikos departamente atlikto gyventojų užimtumo statistinio tyrimo duomenys, siekiant ištirti bedarbių ir užimtųjų gyventojų skaičiaus apskrityse ir savivaldybėse vertinimo tikslumą modeliais grįstais įvertiniais ir informacijos apie asmens registraciją Lietuvos darbo biržoje naudojimo įvertiniuose tikslingumą.

**Reikšminiai žodžiai:** mažo imties dydžio sritis, modeliu pagrįstas įvertinys, daugialypės regresijos modelis, Džeimso ir Štaino įvertinys.

## 1. Įvadas

Įvertinus baigtinės populiacijos parametrus imties planu pagrįstais įvertiniais ir gavus reikiamą įverčių tikslumą, dažnai atsiranda poreikis įvertinti ir populiacijos poaibių, vadinamų sritimis, analogiškus parametrus imties planu pagrįstais įvertiniais. Jei sudarant imties planą sritis nebuvo išskirta į atskirą sluoksnį, tai išrinkus iš populiacijos imtį paprastai gaunasi taip, kad kuo mažesnė sritis, tuo mažiau joje imties elementų, o kartais srityje nėra nei vieno tos imties elemento. Tada srities parametrų arba neįmanoma įvertinti imties planu pagrįstais įvertiniais, arba tie įvertiniai nepakankamai tikslūs. Ieškoma būdų pasinaudoti papildoma informacija, kuri žinoma iš ėmimo sąrašo dar prieš imties išrinkimą visiems populiacijos elementams, ir pritaikyti modelius. Tokie vertinimo metodai vadinami mažų sričių vertinimo metodais (angl. *small area estimation*), t. y. sričių, kurių vidutinis imties dydis per mažas reikiamo tikslumo įverčiams gauti, parametrų vertinimo metodais. Įvertiniams sudaryti taikomi tiesinės regresijos, logistinės regresijos modeliai bei specifiniai mažo imties dydžio sritims būdingi įvertiniai.

Daug mokslinių tyrimų skirta mažų sričių vertinimo metodams, kurių apžvalga pateikta Rao monografijoje [9]. Tai populiari šių dienų statistinių tyrimų sritis, dažnai susijusi su intensyvaus kompiuterio naudojimo galimybėmis, taikoma įvairiems uždaviniams spręsti. Daugelyje šalių atliekami moksliniai tyrimai, ieškant metodų kuo tiksliau įvertinti statistinio gyventojų užimtumo tyrimo baigtinės populiacijos parametrus. Pavyzdžiui, Lenkijos statistikų tyrimai pristatomi Klimanek [4] straipsnyje, kur taikomi mažo imties dydžio sričių hierarchiniais modeliais grįsti įvertiniai, taip pat naudojantys ir erdvinę komponentę. Mūsų straipsnio tikslas – parinkti ir pritaikyti modelius Lietuvos statistikos departamento gyventojų užimtumo statistinio tyrimo duomenims; ištirti, kaip tiksliai galima įvertinti bedarbių ir užimtų gyventojų skaičių srityse, kurias atitinka apskritys ir savivaldybės. Šiame darbe vertinant bedarbių ir užimtų gyventojų skaičių naudojami tokie metodai:

- taikomas nepaslinktas imties planu grįstas įvertinys tyrimo kinamųjų reikšmių sumoms vertinti;
- sudaromi daugialypės regresijos ir logistinės regresijos modeliai ir šiais modeliais pagrįsti įvertiniai užimtųjų ir bedarbių skaičiui vertinti;
- taikomas kompromisinis Džeimso ir Štaino įvertinys, pagrįstas tiek imties planu, tiek ir modeliu, kuris yra pateiktas Rao [9] knygoje, taip pat jau buvo taikytas Krapavickaitės [6] straipsnyje;

– modeliavimo būdu apskaičiuojami empiriniai įvertinių tikslumo matai ir palyginami tarpusavyje.

Vienas iš svarbių nepriklausomų kintamųjų modeliams sudaryti yra asmens registracija Lietuvos darbo biržoje, žinoma visiems populiacijos elementams. Šio kintamojo apibrėžimas skiriasi nuo statistinio tyrimo bedarbio apibrėžimo [10], bet tarp jų yra stipri statistinė priklausomybė. Šis kintamasis kartais įtraukiamas į modelių pagalbą naudojančius įvertinius vertinant bedarbių skaičių kitų šalių, pavyzdžiui, Norvegijos [3], Suomijos [8] oficialiosios statistikos tyrimuose. Šiuo darbu siekiama išsiaiškinti šio kintamojo vaidmenį sudarant įvertinius Lietuvos duomenims.

## 2. Baigtinėje populiacijoje apibrėžto tyrimo kintamojo reikšmių sumos vertinimo būdai

Šiame skyriuje pateikiami imties planu ir modeliais pagrįsti įvertiniai, kurie vėliau taikomi populiacijos sričių parametrams vertinti.

### 2.1. Imties planu pagrįstas įvertinys

Baigtinės populiacijos parametrų vertinimui dažniausiai taikomi imties planu pagrįsti įvertiniai. Juos taikant naudojami tik iš imties elementų gauti duomenys. Vertinant parametrus populiacijos srityse, gali atsitikti taip, kad išrinkus paprastąją atsitiktinę imtį iš populiacijos, kurioje nors jos srityje nebus imties elementų, arba, jei ir būtų srityje imties elementų, šie elementai gali neturėti reikiamų tiriamojo kintamojo reikšmių. Pavyzdžiui, jeigu reikia įvertinti bedarbių skaičių savivaldybėse, tai išrinkus paprastąją atsitiktinę imtį, mažose srityse imtyje gali būti visi dirbantys asmenys, nors populiacijos atitinkamose srityse ir yra bedarbių. Tokiais atvejais sričių parametrai įvertinami labai netiksliai, o nesant imties elementų srityje, tokioje srityje parametrų neįmanoma įvertinti iš viso.

Tegul $U = \{1,2,\ldots,N\}$ – baigtinė populiacija, sudaryta iš $N$ elementų, kurioje apibrėžtas tyrimo kintamasis $y$ su reikšmėmis $y_1, y_2, \ldots, y_N$. Nagrinėjama šio kintamojo reikšmių suma bei vidurkis

$$t_y = \sum_{k=1}^{N} y_k, \qquad \mu_y = t_y/N.$$

Tariama, kad populiacija $U$ suskaidyta į $D$ nesikertančių sričių $U_1, U_2, \ldots, U_D$, kurioms $U = U_1 \cup \ldots \cup U_D$ ir $U_k \cap U_l = \emptyset$, kai $k \neq l$. Populiacijos sričių dydžiai $N_1, N_2, \ldots, N_D$ tenkina lygybę $N_1 + \ldots + N_D = N$. Tyrimo kintamojo $y$ suma ir vidurkis srityje $U_d$ žymimi

$$t_{yd} = \sum_{k \in U_d} y_k, \qquad \mu_{yd} = t_{yd}/N_d, \qquad d = 1, \ldots, D.$$

Tegul $\mathbf{i} \subset U$ yra iš populiacijos išrinkta $n$ dydžio paprastoji atsitiktinė imtis. Ji taip pat suskyla į $D$ nesikertančių poimčių $\mathbf{i}_1, \ldots, \mathbf{i}_D$, $\mathbf{i} = \mathbf{i}_1 \cup \cdots \cup \mathbf{i}_D$ ir $\mathbf{i}_k \cap \mathbf{i}_l = \emptyset$, kai $k \neq l$. Sričių dydžiai $n_1, \ldots, n_D$ tenkina lygybę $n_1 + \cdots + n_D = n$. Taikomas populiacijos sumos įvertinys

$$\hat{t}_y = \frac{N}{n} \sum_{k \in \mathbf{i}} y_k,$$

kuris yra nepaslinktasis (Krapavickaitė ir Plikusas [5]). Populiacijos vidurkio įvertinys $\hat{\mu}_y = \hat{t}_y/N$ taip pat yra nepaslinktasis.

Jeigu imtyje yra sričiai $U_d$ priklausančių elementų: $\mathbf{i}_d \neq \emptyset$, tai galima įvertinti tyrimo kintamojo $y$ srities reikšmių sumą ir vidurkį:

$$\hat{t}_{yd} = \frac{N}{n} \sum_{k \in \mathbf{i}_d} y_k, \qquad \hat{\mu}_{yd} = \hat{t}_{yd}/N_d, \qquad d = 1, \ldots, D. \tag{1}$$

Kadangi imties planu pagrįstais įvertiniais ne visada galima pakankamai tiksliai įvertinti parametrus srityse, toliau siūlomi kiti trys sudėtingesni įvertiniai, kuriems sudaryti naudojama papildoma informacija.

## 2.2. Logistinės regresijos modelis ir juo pagrįstas sumos įvertinys

Toliau laikoma, kad priklausomas kintamasis $y$ yra dvireikšmis, visiems populiacijos elementams įgyjantis reikšmes 0 arba 1. Taip pat laikoma, kad turima superpopuliacija – nepriklausomi atsitiktiniai dydžiai $Y_1, Y_2, \ldots, Y_N$, kurių realizacijos yra populiacijoje $U$ apibrėžto tyrimo kintamojo $y$ reikšmės $y_1, y_2, \ldots, y_N$. Pasinaudojus papildoma informacija apie populiacijos elementus, bus sudarytas modelis tikimybei, kad atsitiktinis dydis $Y_k$ įgytų reikšmę 1. Kintamasis $y$ vienu atveju reiškia asmens buvimą bedarbiu, kitu atveju – buvimą užimtuoju.

Tariama, kad kintamasis $y$, pavyzdžiui, reiškia buvimą bedarbiu ($y_k = 1$, kai asmuo yra bedarbis, ir $y_k = 0$, kai jis tokiu nėra). Su šia kintamojo reikšme susiejamas dvireikšmis atsitiktinis dydis $Y_k$. Galima sakyti, kad išrinkus imtį stebimos atsitiktinių dydžių $Y_k$, $k = 1,2,...,N$, reikšmės. Tikimybė atsitiktiniam dydžiui $Y_k$ įgyti vienetą priklauso nuo kitų fiksuotų kintamųjų (pavyzdžiui, lyties, amžiaus, gyvenamosios vietos ir pan.), kiekvienam asmeniui nurodomų vektoriumi $\mathbf{x}_k = (x_{k0}, x_{k1}, \ldots, x_{km})'$, čia $x_{k0} = 1$. Apibrėžiama sąlyginė tikimybė, kad $Y_k$ įgis reikšmę 1, jei žinomos papildomų kintamųjų vektoriaus $\mathbf{x}_k$ reikšmės:

$$q(\mathbf{x}_k) = P\{Y_k = 1 | \mathbf{x}_k\}, \qquad k = 1, \ldots, N. \tag{2}$$

Šiai tikimybei įvertinti sudaromas logistinės regresijos modelis, kuris pateiktas Agresti [1], Levulienės [7] knygose:

$$\text{logit}(\mathbf{x}_k) = \ln \frac{q(\mathbf{x}_k)}{1 - q(\mathbf{x}_k)} = \beta_0 + \beta_1 x_{k1} + \cdots + \beta_m x_{km} = \beta' \mathbf{x}_k, \tag{3}$$

arba

$$q(\mathbf{x}_k) = \frac{e^{\beta' \mathbf{x}_k}}{1 + e^{\beta' \mathbf{x}_k}}, \qquad k = 1, \ldots, N. \tag{4}$$

Funkcija $\text{logit}(\mathbf{x}_k)$ apibrėžta srityje $R^{m+1}$, funkcija $q(\mathbf{x}_k)$ su bet kuriuo realiųjų komponenčių vektoriumi $\beta$ įgyja reikšmes iš intervalo $(0,1)$.

Logistinės regresijos modelio koeficientai $\beta = (\beta_0, \beta_1, \ldots, \beta_m)'$ įvertinami didžiausiojo tikėtinumo metodu, įverčių vektorius pažymimas $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_m)'$. Įrašius koeficientus $\widehat{\beta}$ į (4) lygybę, kiekvienam populiacijos elementui galima užrašyti tikimybės $q(\mathbf{x}_k)$ įvertinį:

$$\widehat{q}(\mathbf{x}_k) = \frac{e^{\hat{\beta}' \mathbf{x}_k}}{1 + e^{\hat{\beta}' \mathbf{x}_k}}, \qquad k = 1, \ldots, N. \tag{5}$$

Iš čia sudaromas logistinės regresijos modeliu pagrįstas sumos $t_y$ įvertinys:

$$\hat{t}_y^{log} = \sum_{k=1}^{N} \widehat{q}(x_k). \tag{6}$$

Suskaidžius populiaciją į $D$ nesikertančių sričių, logistinės regresijos modeliu pagrįstas sumos įvertinys srityse atrodo taip:

$$\hat{t}_{yd}^{log} = \sum_{k \in U_d} \widehat{q}(x_k), \qquad d = 1, \ldots, D. \tag{7}$$

Išrinkus paprastąją atsitiktinę imtį ir pasinaudojant visiems populiacijos elementams žinoma papildoma informacija įvertinus $\beta$ koeficientus, nesunkiai gaunamos visų populiacijos elementų tikimybės, kad $Y_k$ įgis reikšmę 1. Kadangi taikant modelį (5) įvertinamos visų populiacijos elementų tikimybės, tai galima gauti įverčius visose populiacijos srityse.

## 2.3. Daugialypės regresijos modelis sritims ir jo pagalba gaunamas įvertinys

Daugialypės regresijos modeliu pagrįstas įvertinys – dar vienas sumos įverinys, kuris gaunamas naudojant papildomą informaciją. Laikant, kad baigtinės populiacijos tyrimo kintamojo $y$ reikšmės yra superpopuliacijos nepriklausomos realizacijos, įvertinys $\hat{\mu}_{yd}$ iš (1) taip pat bus atsitiktinio dydžio realizacija. Daugialypės regresijos

modelis sudaromas nepaslinktiesiems sričių sumų įvertiniams $\hat{\mu}_{yd}$, kuriems kaip papildomi kintamieji naudojama visų sričių agreguota ir iš ėmimo sąrašo žinoma papildoma informacija, pavyzdžiui, lyties, amžiaus, gyvenamosios vietos srities gyventojų dalis. Pirmiausia apskaičiuojami imties planu pagrįsti tyrimo kintamojo vidurkių $\mu_{yd}$ įverčiai $\widehat{\mu}_{yd}$ srityse $U_d$. Po to sudaromas tiesinės regresijos modelis imties planu pagrįstam srities įvertiniui:

$$\widehat{\mu}_{yd} = \gamma_0 + \gamma_1 \mu_{x1d} + \ldots + \gamma_m \mu_{xmd} + \nu_d, \qquad d = 1, \ldots, D, \tag{8}$$

čia $\mu_{x1d}, \ldots, \mu_{xmd}$ – papildomų kintamųjų $x_1, x_2, \ldots, x_m$ sričių $U_d$ vidurkiai, $\nu_d \sim N(0, \sigma_\nu^2)$, $d = 1, 2, ..., D$, – atsitiktinės paklaidos, t. y. nepriklausomi, vienodai pagal normalųjį dėsnį pasiskirstę atsitiktiniai dydžiai su nuliniais vidurkiais ir lygiomis dispersijomis $\sigma_\nu^2$. Mažiausiųjų kvadratų metodu gaunami modelio koeficientų $\gamma_0, \gamma_1, ..., \gamma_m$ įverčiai $\widehat{\gamma}_0, \widehat{\gamma}_1, ..., \widehat{\gamma}_m$, kuriuos įrašius į (8) lygybės dešiniąją pusę, gaunamos tyrimo kintamojo $y$ sričių vidurkių prognozės:

$$\widehat{\mu}^0_{yd} = \widehat{\gamma}_0 + \widehat{\gamma}_1 \mu_{x1d} + \ldots + \widehat{\gamma}_m \mu_{xmd}, \qquad d = 1, \ldots, D. \tag{9}$$

Iš čia gaunami regresiniu modeliu pagrįsti tyrimo kintamojo reikšmių sričių sumų įvertiniai

$$\hat{t}^{reg}_{yd} = N_d \widehat{\mu}^0_{yd}. \tag{10}$$

Srities vidurkių prognozės bus naudojamos kitame skyriuje sudėtingesniems įvertiniams sudaryti.

## 2.4. Džeimso ir Štaino įvertinys

Džeimso ir Štaino įvertinys ([9] 4.4.2 skyrius) sudaromas imties planu pagrįsto nepaslinktojo įvertinio $\widehat{\mu}_{yd}$, apibrėžto (1) formule ir daugialypės regresijos modeliu pagrįstos $\mu_{yd}$ prognozės $\widehat{\mu}^0_{yd}$ srityje $d$ (10) tiesine kombinacija:

$$\widehat{\mu}^\star_{yd(JS)} = \phi_{JS} \widehat{\mu}_{yd} + (1 - \phi_{JS}) \widehat{\mu}^0_{yd}, \quad d = 1, \ldots, D. \tag{11}$$

Koeficientas

$$\phi_{JS} = \frac{\sum_{d=1}^{D} VKP(\widehat{\mu}^0_{yd})}{\sum_{d=1}^{D} \left( VKP(\widehat{\mu}_{yd}) + VKP(\widehat{\mu}^0_{yd}) \right)}$$

parinktas apytiksliai taip, kad minimizuotų sričių vidurkių įvertinių (11) vidutinių kvadratinių paklaidų ($VKP$) imties plano atžvilgiu sumą $\sum_{d=1}^{D} VKP(\widehat{\mu}^\star_{yd(JS)})$. Aišku, kad $\phi_{JS} \in [0, 1]$, bet jis yra nežinomas. Siekiant jį įvertinti daroma prielaida, kad imties planu pagrįsti populiacijos sričių vidurkių įvertiniai $\widehat{\mu}_{yd} = \mu_{yd} + e_d$, $d = 1, ..., D$, yra nepriklausomi atsitiktiniai dydžiai, pasiskirstę pagal normalųjį dėsnį su vidurkiais $\mu_{yd}$ ir lygiomis dispersijomis $D\widehat{\mu}_{yd} = \psi$ (ir $e_d \sim N(0, \psi)$). Daugiklis $\phi_{JS}$, kai $m < D - 2$, vertinamas taip: $\hat{\phi}_{JS} = 1 - \widehat{\psi}/\widehat{\sigma}_v^2$. Čia $\widehat{\sigma}_v^2$ yra daugialypės regresijos modelio (8), laikant $\widehat{\mu}_{yd}$, $d = 1, ..., D$, fiksuotais, liekanų dispersijos $\sigma_v^2$ įvertinys: $\widehat{\sigma}_v^2 = \sum_{d=1}^{D} (\widehat{\mu}_{yd} - \widehat{\mu}^0_{yd})^2 / (D - 2 - m)$. Imties planu pagrįstų įvertinių dispersija $\psi$ vertinama taip:

$$\widehat{\psi} = \frac{1}{n - D} \sum_{d=1}^{D} \sum_{j=1}^{n_d} (y_{dj} - \bar{y}_d)^2 \cdot \frac{1}{D} \sum_{l=1}^{D} \frac{1}{n_l}.$$

Šioje išraiškoje $y_{dj}$ yra $d$-tosios srities $j$-tojo elemento tyrimo kintamojo $y$ reikšmė, $\bar{y}_d = \frac{1}{n_d} \sum_{k \in i_d} y_{dj}$ yra to kintamojo $d$-tosios srities imties vidurkis. Daroma prielaida, kad $0 \le \hat{\phi}_{JS} \le 1$ ir įrašius $\hat{\phi}_{JS}$ į (11) gaunamas Džeimso ir Štaino įvertinys:

$$\widehat{\mu}_{yd(JS)} = \hat{\phi}_{JS} \widehat{\mu}_{yd} + (1 - \hat{\phi}_{JS}) \widehat{\mu}^0_{yd}, \quad d = 1, \ldots, D. \tag{12}$$

Paprastai iš mažo dydžio imties apskaičiuoti imties planu grįsti įvertiniai nors ir yra nepaslinktieji, bet jų dispersijos būna didelės. Modeliu pagrįsti įvertiniai, pavyzdžiui, (9), nors turi nedideles dispersijas, dažniausiai yra paslinktieji. Imamos tiesinės šių įvertinių kombinacijos ir sudaromi kompromisiniai įvertiniai. Toks yra Džeimso ir Štaino (12) įvertinys. Jame imties planu pagrįsti įvertiniai $\widehat{\mu}_{yd}$ yra „pastumiami“ link modeliu pagrįstų įvertinių $\widehat{\mu}^0_{yd}$, naudojant nuo srities nepriklausančius atsitiktinius koeficientus $\hat{\phi}_{JS}$ ir $1 - \hat{\phi}_{JS}$. Jei nepaslinktųjų įvertinių dispersijoms $\widehat{\psi} \sim \widehat{\sigma}_\nu^2$, tai reiškia, kad modeliu pagrįsti įvertiniai nėra labai nutolę nuo imties planu pagrįstų įvertinių, ir koeficientas $1 - \hat{\phi}_{JS}$ prie modeliu pagrįsto įvertinio $\widehat{\mu}^0_{yd}$ Džeimso ir Štaino įvertinyje yra didelis. Bet jei $\widehat{\psi} << \widehat{\sigma}_\nu^2$, kas rodo didelį modelio nuokrypį nuo imties planu pagrįstų įvertinių, tai didesnis

koeficientas $\hat{\phi}_{JS}$ lieka prie imties planu pagrįsto įvertinio $\hat{\mu}_{yd}$. Taip siekiama balanso tarp imties planu ir modeliu pagrįstų įvertinių savybių.

Iš (12) gaunami sričių sumų įvertiniai

$$\hat{t}_{yd(JS)} = N_d \widehat{\mu}_{yd(JS)}, \qquad d = 1,2,...,D. \tag{13}$$

Toliau pateikiami metodai siūlomiems įvertiniams tarpusavyje palyginti.

## 2.5. Įvertinių tikslumo matai

Įvertinus parametrus keliais skirtingais metodais ir norint juos palyginti, reikia pasirinkti tikslumo matus. Įverčių tikslumui palyginti naudojama empirinė dispersija, variacijos koeficientas, poslinkis, vidutinė kvadratinė paklaida ir santykinė vidutinė kvadratinė paklaida.

Iš baigtinės populiacijos pagal tą patį imties planą išrenkama $B$ imčių ir iš kiekvienos jų apskaičiuojama $B$ įvertinio reikšmių $\hat{t}_y^{(i)}$, $i = 1,\ldots,B$. Toliau nurodomi naudojamų tikslumo matų apibrėžimai.

Įvertinio empirinis vidurkis, empirinė dispersija ir variacijos koeficiento įvertinys:

$$\bar{\hat{t}}_y = \frac{1}{B}\sum_{i=1}^{B} \hat{t}_y^{(i)}, \quad \widehat{D}(\hat{t}_y) = \frac{1}{B}\sum_{i=1}^{B} (\hat{t}_y^{(i)} - \bar{\hat{t}}_y)^2, \quad \widehat{cv}(\hat{t}_y) = \frac{\sqrt{\widehat{D}(\hat{t}_y)}}{\bar{\hat{t}}_y}, \tag{14}$$

empirins poslinkis:

$$POSL(\hat{t}_y) = \bar{\hat{t}}_y - t_y, \tag{15}$$

vidutinės kvadratinės paklaidos ($VKP$) įvertinys ir santykinės vidutinės kvadratinės paklaidos ($SVKP$) įvertinys:

$$\widehat{VKP}(\hat{t}_y) = \widehat{D}(\hat{t}_y) + POSL^2(\hat{t}_y), \quad \widehat{SVKP}(\hat{t}_y) = \frac{\sqrt{\widehat{VKP}(\hat{t}_y)}}{\bar{\hat{t}}_y}. \tag{16}$$

Kitame skyriuje pateiktiami metodai bus pritaikyti Lietuvos statistikos departamento gyventojų užimtumo statistinio tyrimo duomenims, esantiems ([11]).

# 3. Statistinis modeliavimas

Antrajame skyriuje buvo trumpai aprašyta teorija, kuri taikoma statistiniam modeliavimui. Toliau aprašomi duomenys ir pateikiami statistinio modeliavimo rezultatai. Visi skaičiavimai atlikti naudojantis statistinės analizės programa *SAS*, o grafikai nubraižyti pasinaudojus programa *R*. Atliekant skaičiavimus vertinamas užimtųjų ir bedarbių skaičius apskrityse ir savivaldybėse.

## 3.1. Pradinių statistinių duomenų aprašymas

Darbe naudojami Lietuvos statistikos departamento gyventojų užimtumo statistinio tyrimo duomenys [11], kurie surinkti iš populiacijos sąrašo išrinkus paprastąją atsitiktinę asmenų imtį ir apklausus visus ne jaunesnius negu 15 metų į imtį išrinktų asmenų namų ūkių narius. Šiame darbe apsiribota ne vyresniais negu 70 metų asmenimis. Statistinio tyrimo duomenys laikomi statistinio modeliavimo populiacija.

Taigi tiriamos modeliavimo populiacijos $U$ dydis $N = 10412$ stebinių, 768 bedarbiai ir 5959 užimtieji. Populiacija suskaidoma į $D$ nesikertančių sričių $U_1, U_2, \ldots, U_D$, atitinkančių apskritis arba savivaldybes. Vienu atveju sričių skaičius yra $D = 10$, nes Lietuvoje yra 10 apskričių, kitu atveju $D = 60$, nes tiek Lietuvoje savivaldybių.

Nagrinėjami du dvireikšmiai tyrimo kintamieji: *bedarbis*, kuris įgyja reikšmę 1, jei asmuo yra bedarbis pagal statistinio tyrimo apibrėžimą, ir 0, jei taip nėra; *užimtasis*, kuris įgyja reikšmę 1, jei asmuo yra užimtasis pagal statistinio tyrimo apibrėžimą, ir 0, jei taip nėra. Jų populiacijos sumos žinomos, todėl išrinkus imtį ir apskaičiavus įverčius, juos galima lyginti su tikrosiomis reikšmėmis.

Modeliavimo populiacijoje yra keletas papildomų kintamųjų, kurių reikšmės žinomos visiems šios populiacijos elementams. Asmens *amžius* suskaidytas į keletą grupių: 1 amžiaus grupė – $15-25$ metų, 2 amžiaus grupė – $26-40$ metų, 3 amžiaus grupė – $41-55$ metų, 4 amžiaus grupė – $56-70$ metų. Kiti kintamieji: *lytis*, gyvenamoji vietovė *miestas / kaimas*, namų ūkio dydis *nu_dydis*, kurio reikšmė lygi 1, jei statistinio tyrimo ėmimo sąraše namų ūkio narių daugiau negu du, ir reikšmė lygi 0 priešingu atveju.

Darbe naudojamas dar vienas papildomas kintamasis *r_d_b*, kuris nurodo, ar tiriamuoju laikotarpiu asmuo buvo registruotas Lietuvos darbo biržoje. Šis kintamasis apibrėžtas kitaip negu statistinio tyrimo *bedarbis*. Šio darbo tikslas – ištirti, ar šio kintamojo populiacijos elementų reikšmių žinojimas gali padėti tiksliau įvertinti bedarbių (pagal statistinio tyrimo apibrėžimą) skaičių populiacijoje ir jos srityse.

Sudaromi dvireikšmiai darbiniai kintamieji *amz_gr1, amz_gr2, amz_gr3, amz_gr4, vyras, miestas*, *nu_dydis*, *r_d_b*. Kiekvienas iš šių kinatmųjų įgyja reikšmę 1, jei asmuo pasižymi atitinkama savybe ir reikšmę 0 priešingu atveju.

## 3.2. Bedarbių ir užimtų gyventojų skaičiaus vertinimas taikant imties planu pagrįstą įvertinį

Imties planu grįsti įvertiniai yra nepaslinktieji. Todėl daug kartų vertinant parametrą pagal tą patį imties planą, gautų įverčių grupių vidurkis yra artimas tikrajai reikšmei. Tačiau vertinamą sritį suskaidžius į žymiai mažesnes, nesikertančias sritis, įverčių tikslumas sumažėja.

Bedarbių ir užimtųjų skaičiaus (sumos) įverčiams srityse gauti renkama paprastoji atsitiktinė imtis (PAI). Visoje populiacijoje ir apskrityse parametrus visada pavyksta įvertinti. Tačiau vertinant parametrus savivaldybėse atsiranda kelios mažos sritys, kurių elementų ne visada būna imtyje. Pavyzdžiui, kartojant imties rinkimą 100 kartų, 32 kartus imtyje nebuvo nė vieno elemento iš Birštono savivaldybės. Variacijos koeficiento įvertis, mažėjant srities imties dydžiui, didėja. Tai reiškia, kad tikslumas mažėja. Užimtųjų skaičiaus įverčiai gauti tiksliau, negu bedarbių skaičiaus įverčiai, nes populiacijoje, o tuo pačiu ir imtyje, vidutiniškai žymiai daugiau užimtųjų nei bedarbių.

## 3.3. Bedarbių ir užimtų gyventojų skaičiaus vertinimas taikant logistinės regresijos modelį

Šiame, o taip pat 3.4 skyriuje pateikiami modeliai su parametrais, įvertintais iš visos populiacijos duomenų. Juose gerai matyti aiškinamųjų kintamųjų vaidmuo modeliuose. Išrinkus imtį modeliai sudaromi jau imties duomenims ir kiekvienai imčiai gaunamos vis kitos modelio koeficientų įverčių reikšmės.

Kadangi kintamieji *bedarbis* ir *užimtasis* yra dvireikšmiai, tai taikant logistinės regresijos modelį modeliuojamos tikimybės, kad atitinkamas kintamasis įgis reikšmę 1.

Pirmiausia logistinės regresijos modeliui sudaryti naudojami visos populiacijos duomenys. Sudarius modelį ir nustačius reikšmingus nepriklausomus kintamuosius, tie kintamieji naudojami sudarant modelį vis naujai išrinktai atsitiktinei imčiai, pagal kurį vertinamos populiacijos tyrimo kintamojo reikšmių sričių sumos.

Naudojant programos *SAS* procedūrą *proc logistic*, gautas toks įvertintas tikimybės būti bedarbiu logistinės regresijos modelis visos populiacijos duomenims:

$$\begin{aligned}\ln\frac{\widehat{q}}{1-\widehat{q}} = &-4{,}06+0{,}42amz\_gr1+0{,}89amz\_gr2+0{,}65amz\_gr3+\\ &-0{,}24miestas+0{,}29vyras+3{,}7r\_d\_b.\end{aligned} \tag{17}$$

Iš koeficientų $\beta_i$ įverčių matoma, kad gyvenant mieste tikimybė būti bedarbiu mažėja, nes prie kintamojo *miestas* koeficiento reikšmė yra neigiama: $\hat{\beta}_4 = -0{,}24$, o užsiregistravimas darbo biržoje tikimybę būti bedarbiu labai padidina, nes koeficiento reikšmė prie kintamojo *r_d_b* yra teigiama ir didelė: $\hat{\beta}_6 = 3{,}7$. Šie $\beta$ koeficientų įverčiai taikant modelį imties duomenims kiekvieną kartą gaunami vis kitokie, bet panašūs.

Kintamojo tikimybei būti užimtuoju visai populiacijai sudarytas logistinės regresijos modelis atrodo taip:

$$\begin{aligned}\ln\frac{\widehat{q}}{1-\widehat{q}} = &-0{,}83-amz\_gr1+2{,}28amz\_gr2+2{,}32amz\_gr3+\\ &+0{,}4miestas+0{,}24vyras-2{,}58r\_d\_b.\end{aligned} \tag{18}$$

Įvertinus koeficientus $\beta_i$ matome, kad, kaip ir buvo galima tikėtis, užsiregistravimas darbo biržoje tikimybę būti užimtuoju sumažina. Didžiausią įtaką tikimybei būti užimtuoju turi priklausymas darbingo amžiaus grupėms – kintamieji *amz_gr*2 ir *amz_gr*3. Jaunimas rečiau dirba, nes dažniausiai studijuoja (neigiamas koeficientas prie *amz_gr*1). Sumoms įvertinti taikoma (7) formulė.

## 3.4. Bedarbių ir užimtųjų skaičiaus vertinimas pagal daugialypės regresijos modelį

### Bedarbių dalies vertinimas taikant daugialypės regresijos modelį

Toliau sudaromi tyrimo kintamųjų vidurkių daugialypės regresijos modeliai sritims. Papildomi kintamieji – gyventojų grupės srities dalis: miesto gyventojų dalis *mies_dalis*, amžiaus grupės dalis *amz*1*_dalis*, *amz*2*_dalis*, *amz*3*_dalis*, didesnių namų ūkių dalis *nu_dyd_dalis*, darbo biržoje registruotų gyventojų dalis *r_d_b*. Apatinis indeksas prie minėtų kintamųjų $d$ šiame skyriuje reiškia jų reikšmes srityje $d$.

Įvertintas daugialypės regresijos modelis apskričių bedarbių dalies įverčiui:

$$\widehat{\mu}^{a}_{bd} = 2{,}56 - 0{,}65 mies\_dalis_d - 0{,}84 amz1\_dalis_d - 1{,}77 nu\_dyd\_dalis_d - 5{,}97 r\_d\_b\_dalis_d. \tag{19}$$

Kaip matome, apskrities bedarbių dalį mažina didesnė miesto gyventojų, jaunimo dalis ir didesnis darbo biržoje užsiregistravusių bedarbių skaičius. Tai reiškia, kad dažnai registruoti darbo biržoje bedarbiai užsiima kažkokia veikla, kuri pagal statistinio tyrimo apibrėžimą juos apibrėžia nebe kaip bedarbius.

Įvertintas daugialypės regresijos modelis savivaldybių bedarbių dalies įverčiui:

$$\widehat{\mu}^{s}_{bd} = 0{,}21 - 0{,}32 amz1\_dalis_d - 0{,}27 amz2\_dalis_d - 0{,}23 amz3\_dalis_d + 0{,}68 r\_d\_b\_dalis_d. \tag{20}$$

Kaip matome, statistinio tyrimo bedarbių dalis savivaldybėse tiesiogiai proporcinga registruotų bedarbių savivaldybės daliai ir atvirkščiai proporcinga darbingo amžiaus savivaldybės gyventojų daliai.

### Užimtų gyventojų dalies vertinimas taikant daugialypės regresijos modelį

Įvertintas daugialypės regresijos modelis apskričių užimtų gyventojų dalies įverčiui:

$$\widehat{\mu}^{a}_{ud} = -0{,}13 - 1{,}33 amz2\_dalis_d + 0{,}44 miestas\_dalis_d + 1{,}58 r\_d\_b\_dalis_d. \tag{21}$$

Šis modelis rodo, kad kurioje apskrityje yra didesnė registruotų bedarbių dalis, ten ir užimtųjų dalis didesnė. Reiškia, mažėja apskrities neaktyvių gyventojų dalis.

Įvertintas daugialypės regresijos modelis savivaldybių užimtų gyventojų dalies įverčiui:

$$\widehat{\mu}^{s}_{ud} = 0{,}27 + 0{,}66 amz2\_dalis_d - 0{,}59 amz3\_dalis_d - 0{,}71 r\_d\_b\_dalis_d. \tag{22}$$

Didesnė savivaldybės registruotų bedarbių dalis mažėjimo kryptimi veikia tos savivaldybės užimtų gyventojų dalį.

### Džeimso ir Štaino įvertinio taikymas

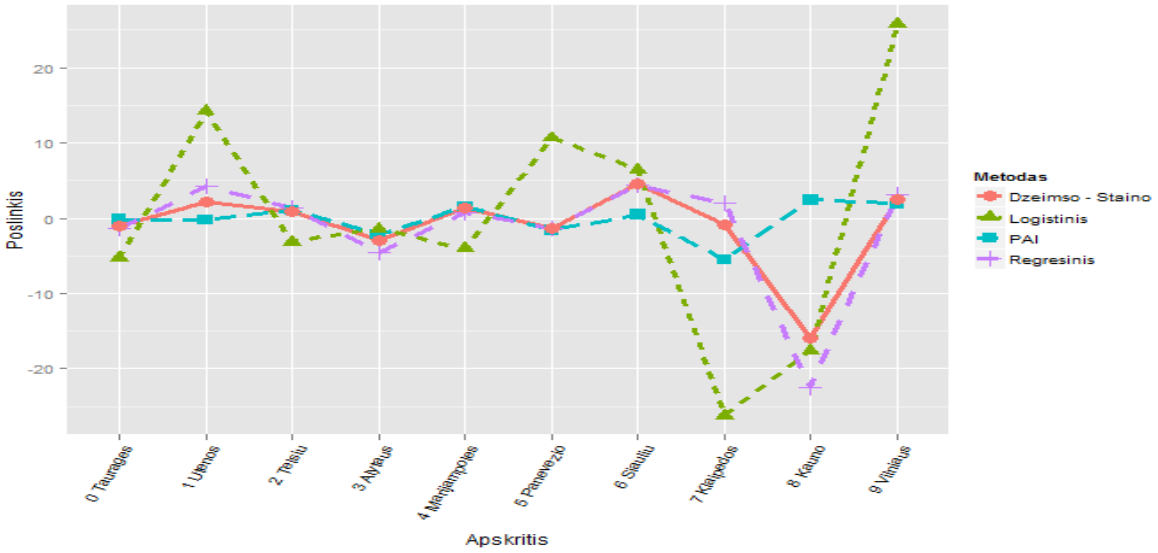
Tyrimo kintamojo srities vidurkio Džeimso ir Štaino įvertinys $\widehat{\mu}_{yd(JS)}$ pateiktas (12) formulėje, padarius prielaidą, kad $0 \le 1 - \hat{\phi}_{JS} \le 1$. Deja, ši prielaida pagrįsta aproksimacijomis ([9]), ir kartais gali negalioti. Jei kuriai nors imčiai koeficientas prie regresijos modeliu pagrįsto įvertinio $1 - \hat{\phi}_{JS} > 1$, tai šis koeficientas keičiamas į 1 ir gaunama $\widehat{\mu}_{yd(JS)} = \widehat{\mu}^{0}_{yd}$. Jei atsitinka, kad į imtį patekusių srities elementų tarpe nėra tokių, kuriems tyrimo kintamasis įgyja reikšmę 1 ir negalima sudaryti įvertinio $\widehat{\mu}_{yd}$, tai imama $\widehat{\mu}_{yd(JS)} = \widehat{\mu}^{0}_{yd}$.

Įrašius įvertinius (19)–(22) į (12) gaunami sričių vidurkių įvertiniai $\widehat{\mu}_{yd(JS)}$, o iš jų pagal (13) – atitinkami sumų įvertiniai.

## 3.5. Bedarbių ir užimtų gyventojų skaičiaus įverčių palyginimas

Kiekvienos iš $B$ imčių atveju modelių parametrai vertinami iš tos imties duomenų ir apskaičiuojami atitinkami populiacijos sričių parametrų įverčiai.

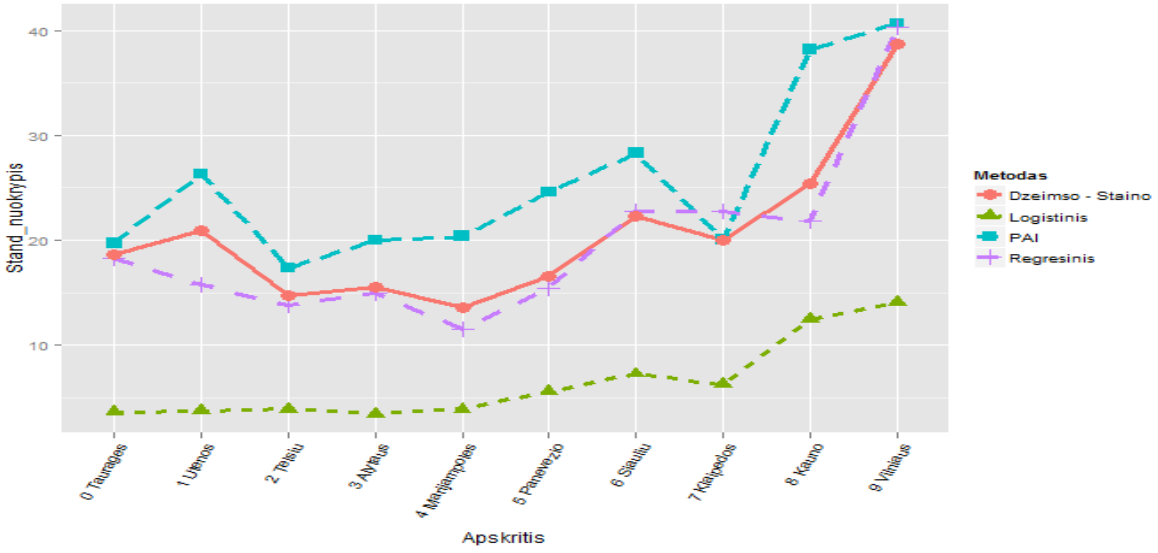
Norint palyginti gautų įverčių tikslumą, buvo apskaičiuoti kiekvienu metodu gautų įverčių empiriniai standartiniai nuokrypiai ir poslinkiai. Iš jų gauti kiti išvestiniai tikslumo matai.




1 pav.: Bedarbių skaičiaus apskrityse įverčių empiriniai poslinkiai

Paveikslėliuose pateikti bedarbių skaičiaus apskrityse įverčių empiriniai tikslumo matai. Apskritys išrikiuotos populiacijos dydžio, o tuo pačiu ir vidutinio imties dydžio didėjimo tvarka.

1 pav. pavaizduotas bedarbių skaičiaus apskrityse įverčių empirinis poslinkis. Matome, kad logistinės regresijos modeliu pagrįsto įvertinio empirinis poslinkis yra didžiausias. Mažiausią empirinį poslinkį turi imties planu pagrįstas įvertinys. Džeimso ir Štaino įvertinio poslinkis mažesnis negu regresiniu modeliu pagrįsto įvertinio, bet didesnis negu imties planu pagrįsto įvertinio.




2 pav.: Bedarbių skaičiaus apskrityse įverčių empiriniai standartiniai nuokrypiai

Bedarbių skaičiaus įverčių empiriniai standartiniai nuokrypiai pavaizduoti 2 pav. Matome, kad mažiausias empirinis standartinis nuokrypis gautas logistinės regresijos metodu. Didžiausias empirinis standartinis

nuokrypis gautas imties planu pagrįstam įvertiniui.

Mažų sričių įvertinių esmė yra ta, kad juose naudojama ne tiktai vertinamos srities informacija, bet ir kitų sričių informacija. Šie įvertiniai gali pagerinti įverčių tikslumą srityse, bet nebūtinai visose srityse. Todėl metodams lyginti imami visų sričių tikslumo matų vidurkiai, ir tiksliausiu laikomas tas įvertinys, kuriam toks matų vidurkis yra mažiausias. Rezultatai pateikiami 1 lentelėje. Regresiniu modeliu pagrįsto įvertinio $SVKP$ svyruoja tarp imties planu pagrįsto ir logistinės regresijos modeliu pagrįsto įvertinio $SVKP$, būdama šiek tiek didesnė negu Džeimso ir Štaino įvertinio $SVKP$.

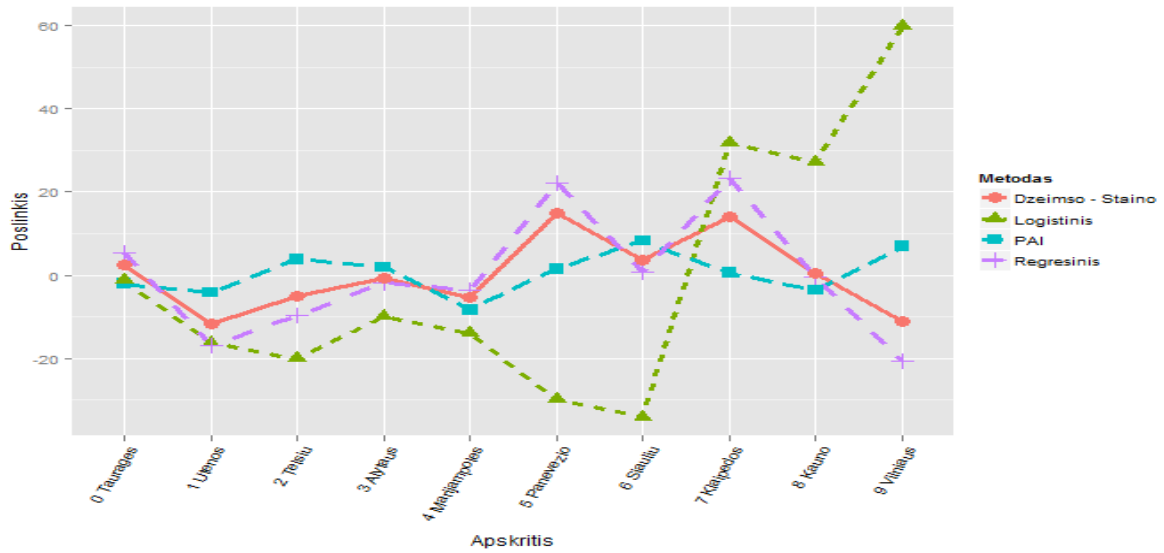
1 lentelė: Bedarbių skaičiaus apskrityse įverčių empirinių tikslumo matų vidurkiai

| Metodas | Stand. nuokrypis | Variacijos koeficientas | $VKP$ | Poslinkis | $SVKP$ |
|---|---|---|---|---|---|
| PAI | 25,549 | 0,383 | 716,236 | −0,222 | 0,385 |
| Regresinis | 19,755 | 0,300 | 509,141 | −1,370 | 0,310 |
| Džeimso ir Štaino | 20,643 | 0,313 | 504,934 | −1,079 | 0,318 |
| Logistinis | 6,438 | 0,084 | 261,880 | −0,074 | 0,176 |

Modeliavimo populiacijos bedarbių skaičiaus apskričių vidurkis lygus 77. 1 lentelėje matyti, kad logistinės regresijos modeliu pagrįsto įvertinio visi tikslumo matai yra mažiausi. Kitiems įvertiniams esant mažam įvertinio poslinkiui padidėja jo standartinis nuokrypis, o mažėjant standartiniam nuokrypiui, išauga įvertinio poslinkis. Ir į poslinkį, ir į standartinį nuokrypį atsižvelgiama vidutinėje kvadratinėje paklaidoje, o dar geriau – santykinėje vidutinėje kvadratinėje paklaidoje, kuri nepriklauso nuo mato vienetų.

Taigi mažiausia santykinė vidutinė kvadratinė paklaida gauta logistinės regresijos modeliu pagrįsto bedarbių skaičiaus apskrityse įvertinio atveju. Džeimso ir Štaino įvertinio $SVKP$ yra tarp imties planu pagrįsto ir regresiniu modeliu pagrįsto įvertinio $SVKP$.

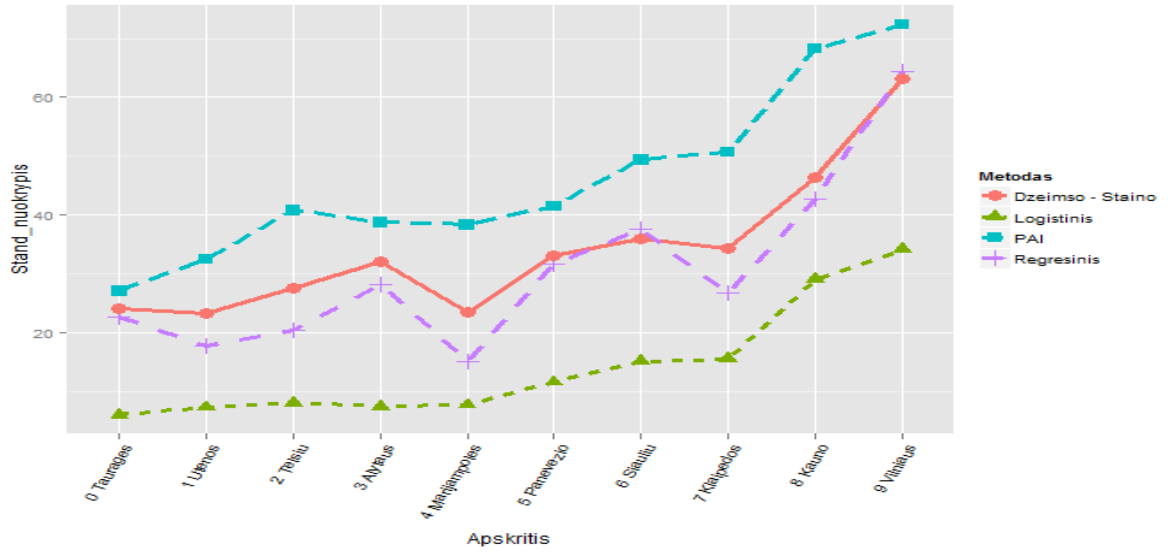
Toliau lyginami užimtųjų skaičiaus įverčių empiriniai tikslumo matai.




3 pav.: Užimtųjų skaičiaus apskrityse įverčių empiriniai poslinkiai

3 pav. matyti, kad šiek tiek didesnį poslinkį, negu pagrįstas imties planu, turi daugialypės regresijos modeliu pagrįstas įvertinys, o didžiausią – logistinės regresijos modeliu pagrįstas įvertinys.

4 pav. matyti užimtų gyventojų skaičiaus įverčių empiriniai standartiniai nuokrypiai apskrityse. Didžiausi empriniai standartiniai nuokrypiai gauti imties planu pagrįstu įvertiniu, o mažiausi – logistinės regresijos modeliu pagrįstu įvertiniu.




4 pav.: Užimtųjų skaičiaus apskrityse įverčių empiriniai standartiniai nuokrypiai

Tiksliausiam užimtų gyventojų skaičiaus vertinimo metodui nustatyti skaičiuojami užimtųjų skaičiaus apskrityse įverčių empirinių tikslumo matų vidurkiai, kurie pateikti 2 lentelėje.

2 lentelė: Užimtųjų skaičiaus apskrityse įverčių empirinių tikslumo matų vidurkiai

| Metodas | Stand. nuokrypis | Variacijos koeficientas | $VKP$ | Poslinkis | $SVKP$ |
|---|---|---|---|---|---|
| PAI | 45,937 | 0,097 | 2327,313 | 0,478 | 0,098 |
| Regresinis | 30,683 | 0,062 | 1324,581 | −0,140 | 0,069 |
| Džeimso ir Štaino | 34,302 | 0,072 | 1390,107 | 0,176 | 0,075 |
| Logistinis | 14,146 | 0,024 | 1119,596 | −0,649 | 0,050 |

Modeliavimo populiacijos užimtų gyventojų skaičiaus apskričių vidurkis lygus 596. Nors logistinės regresijos modeliu grįsto įvertinio santykinė vidutinė kvadratinė paklaida yra mažiausia, jo empirinis poslinkis labai didelis. Džeimso ir Štaino įvertinio $SVKP$ beveik tokia pat kaip ir regresiniu srities lygio modeliu pagrįsto įvertinio.

Kadangi savivaldybių įverčiams pavaizduoti reikėtų labai didelių grafikų, kad kas nors galėtų juose matytis, tai pateikti tiktai skaitiniai rezultatai 3 bei 4 lentelėse.

3 lentelė: Bedarbių skaičiaus savivaldybėse įverčių empirinių tikslumo matų vidurkiai

| Metodas | Stand. nuokrypis | Variacijos koeficientas | $VKP$ | Poslinkis | $SVKP$ |
|---|---|---|---|---|---|
| PAI | 9,669 | 1,155 | 122,612 | −0,082 | 1,161 |
| Regresinis | 4,082 | 0,305 | 59,148 | −0,187 | 0,454 |
| Džeimso ir Štaino | 9,020 | 0,886 | 106,135 | −0,070 | 0,917 |
| Logistinis | 1,507 | 0,119 | 22,653 | −0,115 | 0,397 |

4 lentelė: Užimtųjų skaičiaus savivaldybėse įverčių empirinių tikslumo matų vidurkiai

| Metodas | Stand. nuokrypis | Variacijos koeficientas | $VKP$ | Poslinkis | $SVKP$ |
|---|---|---|---|---|---|
| PAI | 18,063 | 0,326 | 420,981 | −0,022 | 0,327 |
| Regresinis | 8,234 | 0,119 | 654,701 | 4,679 | 0,157 |
| Džeimso ir Štaino | 16,851 | 0,304 | 364,660 | 0,326 | 0,306 |
| Logistinis | 2,333 | 0,025 | 170,890 | 0,109 | 0,091 |

3 ir 4 lentelių rezultatai rodo tokias pat savivaldybių bedarbių ir užimtų gyventojų skaičiaus įverčių $SVKP$ kitimo tendencijas, kaip ir 1 bei 2 lentelėse, kur pateikiami modeliavimo rezultatai apskritims. Didelį užimtų gyventojų skaičiaus savivaldybėse regresiniu modeliu pagrįsto įvertinio poslinkį 4 lentelėje nulemia labai didelis

vienos (Vilniaus m.) savivaldybės populiacijos ir imties užimtų gyventojų skaičius ir labai didelis regresiniu modeliu grįstų jos įverčių poslinkis, lyginant su kitomis savivaldybėmis.

## 4. Išvados

Darbe išnagrinėti keturi baigtinės populiacijos parametrų vertinimo metodai. Taikant nepaslinktąjį imties planu pagrįstą įvertinį, įsitikinta, kad jo poslinkis mažiausias, tačiau dipsersija didžiausia. Esant tokioms mažoms populiacijos sritims, kaip savivaldybė, išrinkus paprastąją atsitiktinę imtį, kartais būna taip, kad kurioje nors srityje nėra imties elementų, ir tada įvertinti parametrų neįmanoma. Šiam uždaviniui spręsti buvo sudaromi logistinės regresijos modeliu ir daugialypės regresijos modeliu pagrįsti įvertiniai. Tačiau nors tokiais metodais gauti įvertiniai turi mažą dispersiją, dažnai jie turi didesnius poslinkius.

Modeliavimo rezultatai rodo, kad logistinės regresijos modeliu pagrįsto įvertinio $SVKP$ yra mažiausia. Regresiniu modeliu pagrįsto įvertinio $SVKP$ žymiai mažesnė negu imties planu pagrįsto įvertinio. Džeimso ir Štaino įvertinio visi tikslumo matai mažesni negu imties planu pagrįsto įvertinio, bet didesni negu regresiniu modeliu pagrįsto įvertinio.

Visiems modeliams sudaryti buvo reikšmingas registruotos bedarbystės kintamasis. Prie jo esančio koeficiento absoliutinis dydis visada gana didelis.

**Padėka.** Nuoširdžiai dėkojame recenzentams už dalykiškas pastabas, padėjusias patobulinti straipsnį.

# Literatūra

[1] Agresti, A., *Categorical Data Analysis*, Wiley & Sons, 2013. 744 p.

[2] Čekanavičius, V., Murauskas, G., *Statistika ir jos taikymai* II, Vilnius: TEV, 2002. 272 p.

[3] Hamre, J. I., and Heldal, J., *Improved calculation and dissemination of coefficients of variation in the Norwegian LFS*, 2013. Oslo–Kongsvinger: Statistics Norway [interaktyvus], [žiūrėta 2015 m. rugsėjo 30 d.]. Prieiga per internetą: <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/_attachment/148090?_ts=142476a8ad0>.

[4] Klimanek, T. Using indirect estimation with spatial autocorrelation in social surveys in Poland, *Journal of Przegląd Statystyczny*, 2012, 59, p. 155–172.

[5] Krapavickaitė, D., Plikusas, A. *Imčių teorijos pagrindai.* Vilnius: Technika, 2005. 312 p.

[6] Krapavickaitė, D., An example of small area estimation in finite population sampling. *Lietuvos matem. rink.*, 2003, 43(spec.nr.), p. 497–503.

[7] Levulienė, R., *Statistikos taikymai naudojant SAS.* Vilnius: Vilniaus universiteto leidykla, 2009. 364 p.

[8] Quatember, A. *A comparison of the the Labour Force Surveys of the DACSEIS project from a sampling theory point of view*, 2002 [interaktyvus], [žiūrėta 2015 m. rugsėjo 30 d.]. Prieiga per internetą: <https://www.unitrier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/DRPS3.pdf>.

[9] Rao, J. N. K. *Small Area Estimation*. Hoboken: John Wiley & Sons, 2003. 313 p.

[10] Statistikos departamentas. *Gyventojų užimtumo statistinio tyrimo metodika* [interaktyvus]. Vilnius [žiūrėta 2013 m. gruodžio 1 d.]. Prieiga per internetą: <http://www.stat.gov.lt/lt/>.

[11] Statistikos departamentas. *Viešosios duomenų rinkmenos. Ketvirtinio gyventojų užimtumo statistinio tyrimo* [interaktyvus]. Vilnius [žiūrėta 2015 m. rugpjūčio 31 d.]. Prieiga per internetą: <http://osp.stat.gov.lt/viesos-duomenu-rinkmenos>.

# ESTIMATION OF THE NUMBER OF UNEMPLOYED AND EMPLOYED FOR LITHUANIAN COUNTIES AND MUNICIPALITIES USING MODELS

**Linas Naujanis, Danutė Krapavickaitė**

**Abstract**

Problems of finite population parameters estimation are analyzed in this paper. Four methods have been used for parameter estimation: sampling design-based unbiased estimator, multiple regression and logistic regression model-based estimators and James–Stein estimator. The design-based estimator is unbiased, but its standard deviation is usually high. Model-based estimators are not unbiased, but their standard deviations are low. In order to minimize the standard deviation and the bias, the James–Stein estimator is applied. Labour force survey data of Statistics Lithuania are used for simulation to study model-based estimators for the number of unemployed and employed persons in districts and counties, and the role of information on registered unemployment in these models.

**Key words:** area with small sample size, model-based estimator, multiple regression model, James–Stein estimator.