VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Milda Aleknavičiūtė

Saviranko taikymas baigtinės populiacijos dispersijai vertinti

Magistro darbas

VILNIUS 2006

Matematinės analizės katedra

Darbo vadovas **doc. dr. A. Plikusas** ________________________
(parašas)

Darbas apgintas ______*2006 m. birželio 1 d.*______
Gynimo posėdžio protokolo Nr. ___________________

Registravimo NR.____________________
2006-05-20 __________________________

Turinys

1 Santrauka

Šiame darbe yra nagrinėjama baigtinė populiacija iš kurios išrinkta sluoksninė imtis su paprastąja atsitiktine imtimi sluoksniuose. Nagrinėjamas populiacijos parametras yra populiacijos dispersija. Pateikti trys jos įvertiniai ir parodyta, kad du iš jų yra nepaslinktieji, o vienas turi poslinkį ir šis poslinkis yra apskaičiuotas. Taip pat nagrinėjami penki savirankos metodai skirti dispersijos dispersijai vertinti: tradicinė saviranka, mastelio pakeitimo saviranka, veidrodinio sutapimo saviranka, negrąžintinė saviranka ir grąžintinė saviranka. Palyginimui yra naudojama apytikslės dispersijos įvertinio dispersijos formulė specialiai išvesta sluoksninėms imtims. Šio darbo tikslas palyginti savirankos metodais gautas dispersijos įvertinio dispersijos įvertinio reikšmes tarpusavyje ir su reikšmėmis gautomis skaičiuojant pagal apytikslę formulę.

2 Abstract

Consider a finite population from which the stratified sample with simple random sample without replacement in each strata is drawn. The finite population parameter of interest is variance. There are viewed three estimators of variance and are shown that two estimators are unbiased and one estimator has a bias which is estimated. We studied five bootstrap methods, the naive bootstrap, the rescaling bootstrap, the mirror-match bootstrap, the without-replacement bootstrap and the with-replacement bootstrap for variance estimation of estimator of variance. In comparison, there is used approximation variance formula of variance which is derived special for stratified samples. Our purpose was to compare the result of the bootstrap methods with each other and with the result from the approximation variance formula.

3 Įvadas

Pirmas bet kokio statistinio tyrimo žingsnis - pasirinkti tiriamą aibę. Jos objektai turi vieną ar keletą tyrėją dominančių požymių. Pavyzdžiui, sociologą domina pensinio amžiaus žmonių balsavimo prioritetai. Tiriama aibė - pensinio amžiaus žmonės. Požymis, kuris šiame tyrime domina sociologą, - balsavimo prioritetai. Mokesčių inspekciją domina visų uždarųjų akcinių bendrovių metinės pajamos. Tyrimo objektų aibė - visos UAB, požymis - metinės pajamos. Statistinių tyrimų nagrinėjama objektų aibė dar vadinama populiacija.

Paprastai, atliekant realius tyrimus, požymių pasiskirstymo dažniai populiacijoje tiksliai nežinomi. Be to, dažnai visos populiacijos ištirti neįmanoma. Pagrindinės priežastys dėl kurių visa populiacija tiriama retai yra laiko stoka, didelė kaina ir negalėjimas išvardyti visų populiacijos elementų (pavyzdžiui, kai populiacija begalinė). Taigi dažniausiai tiriama tik populiacijos dalis, vadinama imtimi (pavyzdžiui, tūkstantis apklausoje dalyvavusių rinkėjų; dešimt mokesčių inspekcijos patikrintų firmų).

Tyrimai, kai renkami ir nagrinėjami imties elementų duomenys, siekiant padaryti išvadas apie visą populiaciją , vadinami imčių tyrimais. Pavyzdžiui, pensinio amžiaus žmonių balsavimo prioritetų tyrime yra apklausiama nedidelė pensinio amžiaus žmonių dalis ir, remiantis gauta informacija, yra įvertinami visų pensinio amžiaus žmonių balsavimo prioritetai. Imčių teorijoje nagrinėjama, kaip reikia išrinkti imtį ir ja naudojantis įvertinti populiacijos charakteristikas. Populiacijos parametras - tai baigtinės populiacijos skaitinė charakteristika, kuri gali būti randama, kai turime tyrimo kintamojo reikšmes visiems populiacijos elementams. Baigtinės populiacijos parametru gali būti populiacijos suma, vidurkis ir kiti. Taip pat įdomūs ir netiesiniai baigtinės populiacijos parametrai, tokie, kaip dispersija, sumų santykis, koreliacijos koeficientas ir kiti. Nagrinėjamą populiacijos parametrą žymėsime θ, o parametro įvertinį $\widehat{\theta}$. Įvertinys - tai imties duomenų funkcija (formulė), kuria remiantis yra įvertinamas populiacijos parametras.

Svarbią imčių teorijos dalį sudaro populiacijos parametrų įvertinių $\widehat{\theta}$ dispersijų $D\widehat{\theta}$ vertinimas. Dispersijos įvertiniai $\widehat{D\widehat{\theta}}$ naudojami parametrų įvertinių tikslumo nustatymui. Naudojant šiuolaikinę skaičiavimo techniką, galima skaičiuoti dispersijų įverčius net ir esant sudėtingoms įvertinių išraiškoms. Pageidautina, kad dispersijos $D\widehat{\theta}$ įvertiniai $\widehat{D\widehat{\theta}}$ turėtų tokias savybes:

1. Dispersijos įvertinys būtų nepaslinktasis arba beveik toks, t.y. tenkinantis sąlygą

$$E(\widehat{D}\widehat{\theta}) \approx D\widehat{\theta}.$$

2. Stabilus t.y. dispersijos įvertinio dipersija D($\widehat{D}\widehat{\theta}$) turi būti maža.
3. Neneigiamas ($\widehat{D}\widehat{\theta} \geq 0$) bet kuriai galimai imčiai.

Savirankos metodą pirmą kartą pristatė Efron (1979). Jis buvo taikomas imtims su nepriklausomais ir vienodai pasiskirsčiusiais stebėjimais turintiems kažkokį pasiskirstymą F. Nuo to laiko, buvo atlikta daug teorinių ir empyrinių tyrimų nagrinėjančių savirankos įvertinių savybes ir saviranka tapo populiariu įrankiu klasikinėje statistinėje analizėje.

Savirankos metodas veikia tiek glodiems tiek ir neglodiems įvertiniams, jis reikalauja didelio kiekio nustatytų skaičiavimų, kurie paprastai nėra rimta problema turint šių laikų galingus kompiuterius. Savirankos metodo asimptotinį pagrįstumą įrodė Rao ir Wu (1988) glodžioms populiacijos sumos funkcijoms. Panašių rezultatų tikimasi baigtinės populiacijos dispersijos atveju, bet griežtų įrodymų šiuo atveju nėra.

Savirankos metodo idėja - iš pradinės imties pakartotinai renkant poimčius, sukurti tokias duomenų aibes, iš kurių galima įvertinti populiacijos parametrų įvertinių išsibarstymo charakteristikas, nenaudojant didelių analizinių išvedžiojimų. Taikant savirankos metodą, iš pradinės imties daug kartų modeliuojama nauja savirankos imtis ir parametro vertinimo procedūra daug kartų pakartojama.

Šiame darbe yra nagrinėjamos keturios sugeneruotos populiacijos, iš kurių dviejų populiacijų elementai pasiskirstę pagal normalinį dėsnį, o kitų dviejų - pagal eksponentinį dėsnį. Taip pat šiame darbe yra nagrinėjamos dvi ekonominės veiklos (duomenys pateikti priede), kurios sudaro dvi populiacijas. Tokios generuotos populiacijos yra imamos turint mintyje jų praktinę reikšmę, nes daugelis statistinių išvadų remiasi prielaida apie duomenų normalinį pasiskirstymą (pavyzdžiui, šia prielaida remiamasi sudarant pasikliautinumo intervalus), o eksponentinis pasiskirstymas dažnai pasitaiko tarp realių duomenų.

Kiekvienoje populiacijoje bus renkama sluoksninė imtis su paprastąja atsitiktine imtimi sluoksniuose. Paprastoji atsitiktinė imtis arba paprastoji negrąžintinė imtis - tai tokia n skirtingų elementų imtis iš N dydžio baigtinės populiacijos, kai bet kuris n skirtingų elementų rinkinys turi vienodą tikimybę būti išrinktu. Imties planas, kai suskaidžius populiaciją į kelias nesikertančias dalis, vadinamas sluoksniais, renkamos paprastosios atsitiktinės imtys iš kiekvieno sluoksnio atskirai, nepriklausomai nuo kitų, vadinamas sluoksnine imtimi su paprastąja atsitiktine imtimi sluokniuose.

Naudojant aukščiau paminėtą imties planą, kiekvienoje iš populiacijų yra nagrinėjamas jos parametras - populiacijos kintamojo y dispersija S^2 ir jos įvertinys $\widehat{S^2}$, kuris aprašytas sekančiame skyriuje.

Šio darbo tikslas yra taikant sluokninį ėmimą su paprastąja atsitiktine imtimi sluokniuose ir savirankos metodus įvertinti nagrinėjamų populiacijų dispersijas ir dispersijų įvertinių dispersijas. Taip pat palyginti dispersijos įvertinių dispersijos įvertinių reikšmes gautas taikant savirankos metodus tarpusavyje ir su tikrąją dispersijos įvertinio dispersijos įvertinio reikšme, kurią apskaičiuosime pagal žinomą apytikslę dispersijos įvertinio dispersijos formulę. Tuo pačiu siekiame palyginti kaip veikia savirankos metodai skirtingai sugeneruotose populiacijose ir populiacijose su realiais duomenimis.

Teorinę šio darbo dalį sudaro ketvirtas ir penktas skyrius. Ketvirtame skyriuje išdėstoma, kaip yra vertinama populiacijos dispersija, o penktas skyrius yra skirtas savirankos rūšių nagrinėjimui. Šeštame ir septintame skyriuje yra aprašoma šio darbo eksperimentinė dalis ir gauti rezultatai. Aštuntame skyriuje pateiktos išvados.

4 Populiacijos dispersijos įvertiniai

Tarkime, norime sužinoti kintamojo y išsibarstymą apie jo vidutinę reikšmę N dydžio baigtinėje populiacijoje. Tam yra skaičiuojama kintamojo y dispersija S^2 pagal formulę :

$$S^2 = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - \mu)^2,$$

kur

$$\mu = \frac{1}{N}\sum_{k=1}^{N} y_k.$$

Šią formulę galime perrašyti taip:

$$S^2 = \frac{1}{N-1}\sum_{k=1}^{N}(y_k-\mu)^2 = \frac{1}{2N(N-1)}\sum_{k,l=1}^{N}(y_k-y_l)^2 = \frac{1}{N}\sum_{k=1}^{N} y_k^2 - \frac{1}{N(N-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N} y_k y_l$$

Kintamojo y dispersija priklauso nuo visos populiacijos kintamojo y reikšmių ir realiame tyrime mes negalime jos suskaičiuoti. Todėl dispersiją S^2 vertineme konstruodami jos įvertinį $\widehat{S}^2$.

Paprastosios atsitiktinės imties atveju dispersijos įvertinys yra toks:

$$\widehat{S}_1^2 = \frac{N}{(N-1)n}\sum_{k\in s}(y_k - \widehat{\mu})^2,$$

čia

$$\widehat{\mu} = \frac{1}{n}\sum_{k\in s} y_k.$$

Galimi ir tokie dispersijos įvertiniai:

$$\widehat{S}_2^2 = \frac{1}{2n(n-1)}\sum_{k,l\in s}(y_k - y_l)^2$$

$$\widehat{S}_3^2 = \frac{1}{n}\sum_{k\in s} y_k^2 - \frac{1}{n(n-1)}\sum_{\substack{k,l\in s\\k\neq l}} y_k y_l$$

Žymėjimas s įvertiniuose $\widehat{S}_1^2$,$\widehat{S}_2^2$ ir $\widehat{S}_3^2$ reiškia tikimybinę imtį. Įvertiniai $\widehat{S}_2^2$ ir $\widehat{S}_3^2$ yra nepaslinktieji, tačiau įvertinys $\widehat{S}_1^2$ nėra nepaslinktasis.

Sluoksninės imties su paprastąja atsitiktine imtimi sluoksniuose atveju dispersijos įvertinys yra toks:

$$\widehat{S}_1^2 = \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(y_k - \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2$$

Šiuo atveju galimi ir tokie dispersijos įvertiniai:

$$\widehat{S}_2^2 = \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{k,l\in s_h}(y_k - y_l)^2 + \sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f}(y_k - y_l)^2\right)$$

$$\widehat{S}_3^2 = \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k^2 - \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{k,l\in s_h, k\neq l} y_k y_l -$$

$$-\frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f} y_k y_l$$

Įvertiniai $\widehat{S}_2^2$ ir $\widehat{S}_3^2$ yra nepaslinktieji, o įvertinys $\widehat{S}_1^2$ nėra nepaslinktasis (įrodymai skyriaus pabaigoje).

Apytikslis parametro S^2 įvertinio $\widehat{S}^2$ dispersijos $D\widehat{S}^2$ įvertinys $\widehat{D}\widehat{S}^2$ (kuris darbe vadinamas tikrąja reikšme) skaičiuojamas pagal formulę iš knygos [5]:

$$\widehat{D}\widehat{S}^2 = \sum_{k,l\in s}\left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\frac{\widehat{u}_k}{\pi_k}\frac{\widehat{u}_l}{\pi_l}$$

čia

$$\widehat{u}_k = \frac{1}{N-1}\left[\left(y_k - \frac{1}{N}\sum_{k\in s}\frac{y_k}{\pi_k}\right)^2 - \frac{1}{N-1}\sum_{k\in s}\frac{(y_k - \frac{1}{N}\sum_{k\in s}\frac{y_k}{\pi_k})^2}{\pi_k}\right]$$

s-tikimybinė imtis;
$\pi_k = P(k \in s)$-tikimybė, kad elementas k priklauso imčiai s;
$\pi_{kl} = P(k\&l \in s)$-tikimybė, kad elementai k ir l priklauso s;
Sluoksninės imties su paprastąja atsitiktine imtimi sluoksniuose atveju (išvedimas pateiktas skyriaus pabaigoje):

$$\widehat{D}\widehat{S}^2 = \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} \widehat{s}_h^2, \qquad (*)$$

čia

$$\widehat{s}_h^2 = \frac{1}{n_h - 1} \sum_{k \in s_h} (\widehat{u}_{hk} - \widehat{\mu}_{hk})^2$$

$$\widehat{\mu}_{hk} = \frac{1}{n_h} \sum_{k \in s_h} \widehat{u}_{hk}$$

$$\widehat{u}_{hk} = \frac{1}{N-1}\left[\left(y_{hk} - \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 - \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(y_k - \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2\right].$$

1. Teiginys. Dispersijos S^2 įvertinys $\widehat{S}_1^2$ nėra nepaslinktasis.

Įrodymas.

$$\widehat{S}_1^2 = \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(y_k - \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 =$$

$$= \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(y_k^2 - 2y_k\left(\frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right) + \left(\frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2\right) =$$

$$= \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k^2 - 2\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 +$$

$$+ \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(\frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 = \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k^2 -$$

$$-2\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 + \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h} n_h \frac{1}{N^2}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2 =$$

$$= \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k^2 - \frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2.$$

$$E\left(\frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}y_k^2\right)=E\left(\frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2 I_{hk}\right)=\frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2 EI_{hk}=$$

$$=\frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2\frac{n_h}{N_h}=\frac{1}{N-1}\sum_{h=1}^{H}\sum_{k=1}^{N_h}y_k^2=\frac{1}{N-1}\sum_{h=1}^{N}y_k^2.$$

$$E\left(\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}y_k\right)^2\right)=$$

$$E\left(\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\left(\frac{N_h}{n_h}\sum_{k\in s_h}y_k\right)^2+\sum_{\substack{h,f=1\\h\neq f}}^{H}\left(\frac{N_h}{n_h}\sum_{k\in s_h}y_k\right)\left(\frac{N_f}{n_f}\sum_{l\in s_f}y_l\right)\right)\right)=$$

$$=E\left(\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{k\in s_h}y_k^2+\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{\substack{k,l\in s_h\\k\neq l}}y_k y_l+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f}y_k y_l\right)\right)=$$

$$=E\left(\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{k=1}^{N_h}y_k^2 I_{hk}+\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l I_{hk}I_{hl}+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}y_k y_l I_{hk}I_{fl}\right)\right)=$$

$$=\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{k=1}^{N_h}y_k^2 EI_{hk}+\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l EI_{hk}I_{hl}+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}y_k y_l EI_{hk}I_{fl}\right)=$$

$$=\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{k=1}^{N_h}y_k^2\frac{n_h}{N_h}+\sum_{h=1}^{H}\frac{N_h^2}{n_h^2}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l\frac{n_h(n_h-1)}{N_h(N_h-1)}+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}y_k y_l\frac{n_h}{N_h}\frac{n_f}{N_f}\right)=$$

$$=\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2+\sum_{h=1}^{H}\frac{N_h}{n_h}\frac{n_h-1}{N_h-1}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l+\sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}y_k y_l\right)=$$

$$=\frac{1}{N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2+\sum_{h=1}^{H}\frac{N_h}{n_h}\frac{n_h-1}{N_h-1}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l+\sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}y_k y_l+\right.$$

$$\left.+\sum_{h=1}^{H}\sum_{k=1}^{N_h}y_k^2+\sum_{h=1}^{H}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l-\sum_{h=1}^{H}\sum_{k=1}^{N_h}y_k^2-\sum_{h=1}^{H}\sum_{\substack{k,l=1\\k\neq l}}^{N_h}y_k y_l\right)=$$

$$= \frac{1}{N(N-1)}\left(\left(\sum_{h=1}^{H}\sum_{k=1}^{N_h} y_k\right)^2 + \sum_{h=1}^{H}\left(\frac{N_h}{n_h} - 1\right)\sum_{k=1}^{N_h} y_k^2 + \sum_{h=1}^{H}\left(\frac{N_h(n_h-1)}{n_h(N_h-1)} - 1\right)\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l\right) =$$

$$= \frac{1}{N(N-1)}\left(\left(\sum_{k=1}^{N} y_k\right)^2 + \sum_{h=1}^{H}\frac{N_h - n_h}{n_h}\sum_{k=1}^{N_h} y_k^2 + \sum_{h=1}^{H}\frac{n_h - N_h}{n_h(N_h-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l\right) =$$

$$= \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 + \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h - n_h}{n_h(N_h-1)}\left((N_h - 1)\sum_{k=1}^{N_h} y_k^2 - \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l\right) =$$

$$= \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 + \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h - n_h}{n_h(N_h-1)}\left(N_h\sum_{k=1}^{N_h} y_k^2 - \left(\sum_{k=1}^{N_h} y_k^2 + \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l\right)\right) =$$

$$= \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 + \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h - n_h}{n_h(N_h-1)}\left(N_h\sum_{k=1}^{N_h} y_k^2 - \left(\sum_{k=1}^{N_h} y_k\right)^2\right) =$$

$$= \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 + \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{(N_h - n_h)N_h}{n_h(N_h-1)}\left(\sum_{k=1}^{N_h} y_k^2 - \frac{1}{N_h}\left(\sum_{k=1}^{N_h} y_k\right)^2\right).$$

čia

$$I_{hk} = \begin{cases} 1, & \textbf{kai } k \in s_h \\ 0, & \textbf{kai } k \notin s_h \end{cases}$$

$$k = 1, ..., N_h, \quad h = 1, ..., H.$$

Atsitiktiniai dydžiai I_{hk} vadinami priklausymo imčiai indikatoriais.

Taigi įvertinio $\widehat{S}_1^2$ vidurkis yra

$$E\widehat{S}_1^2 = \frac{1}{N-1}\sum_{h=1}^{N} y_k^2 - \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 -$$

$$-\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{(N_h - n_h)N_h}{n_h(N_h-1)}\left(\sum_{k=1}^{N_h} y_k^2 - \frac{1}{N_h}\left(\sum_{k=1}^{N_h} y_k\right)^2\right).$$

Iš čia matome, kad įvertinys $\widehat{S}_1^2$ nėra nepaslinktasis ir jo poslinkis yra toks:

$$Posl(\widehat{S}_1^2) = E\widehat{S}_1^2 - S_1^2 = \frac{1}{N-1}\sum_{h=1}^{N} y_k^2 - \frac{1}{N(N-1)}\left(\sum_{k=1}^{N} y_k\right)^2 -$$

$$-\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{(N_h-n_h)N_h}{n_h(N_h-1)}\left(\sum_{k=1}^{N_h}y_k^2-\frac{1}{N_h}\left(\sum_{k=1}^{N_h}y_k\right)^2\right)-\frac{1}{N-1}\sum_{k=1}^{N}\left(y_k-\frac{1}{N}\sum_{k=1}^{N}y_k\right)^2=$$

$$=\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{(n_h-N_h)N_h}{n_h(N_h-1)}\left(\sum_{k=1}^{N_h}y_k^2-\frac{1}{N_h}\left(\sum_{k=1}^{N_h}y_k\right)^2\right)=$$

$$=\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{(n_h-N_h)N_h}{n_h(N_h-1)}\sum_{k=1}^{N_h}(y_k-\mu_h)^2=\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(n_h-N_h)}{n_h}s_h^2$$

Čia

$$\mu_h=\frac{1}{N_h}\sum_{k=1}^{N_h}y_k$$

$$s_h^2=\frac{1}{N_h-1}\sum_{k=1}^{N_h}(y_k-\mu_h)^2$$

Šį poslinkį galime vertinti taip:
Darome prielaidas, kad
1. $s_h^2 \to c_h, c_h = const, h = 1, ..., H, \quad kai N, N_h, n_h \to \infty$
2. $\max_h\{\frac{N_h-1}{N-1}\} \le a \le 1$
3. Slauoksnių skaičius H yra fiksuotas
Tada

$$-a^2\sum_{h=1}^{H}\frac{s_h^2}{n_h}\le Posl(\widehat{S}_1^2)\le 0$$

arba

$$-(ac)^2\sum_{h=1}^{H}\frac{1}{n_h}\le Posl(\widehat{S}_1^2)\le 0$$

arba

$$-c^2\sum_{h=1}^{H}\frac{1}{n_h}\le Posl(\widehat{S}_1^2)\le 0$$

Paskutinėse dviejose nelygybėse naudojamės prielaida, kad $s_h^2 \le c^2$, čia $c = const$, $c_h \le c$, $h = 1, ..., H$. $\square$

2. Teiginys. Dispersijos S^2 įvertinys $\widehat{S}_2^2$ yra nepaslinktasis.

Įrodymas :

$$\widehat{S}_2^2 = \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{k,l\in s_h}(y_k-y_l)^2+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f}(y_k-y_l)^2\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{k,l\in s_h}(y_k^2-2y_ky_l+y_l^2)+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f}(y_k^2-2y_ky_l+y_l^2)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\left(n_h\sum_{k\in s_h}y_k^2-2\sum_{k\in s_h}y_k\sum_{l\in s_h}y_l+n_h\sum_{l\in s_h}y_l^2\right)+\right.$$

$$\left.+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k\in s_h}y_k^2-2\sum_{k\in s_h}y_k\sum_{l\in s_f}y_l+n_h\sum_{l\in s_f}y_l^2\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\left(2n_h\sum_{k\in s_h}y_k^2-2\left(\sum_{k\in s_h}y_k\right)^2\right)+\right.$$

$$\left.+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k\in s_h}y_k^2-2\sum_{k\in s_h}y_k\sum_{l\in s_f}y_l+n_h\sum_{l\in s_f}y_l^2\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left(n_h\sum_{k\in s_h}y_k^2-\sum_{k\in s_h}y_k^2-\sum_{\substack{k,l\in s_h\\k\neq l}}y_ky_l\right)+\right.$$

$$\left.+\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k\in s_h}y_k^2-2\sum_{k\in s_h}y_k\sum_{l\in s_f}y_l+n_h\sum_{l\in s_f}y_l^2\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left((n_h-1)\sum_{k\in s_h}y_k^2-\sum_{\substack{k,l\in s_h\\k\neq l}}y_ky_l\right)+\right.$$

$$\left.\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k\in s_h}y_k^2-2\sum_{k\in s_h}y_k\sum_{l\in s_f}y_l+n_h\sum_{l\in s_f}y_l^2\right)\right) =$$

$$E\left(\frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left((n_h-1)\sum_{k\in s_h}y_k^2-\sum_{\substack{k,l\in s_h\\k\neq l}}y_ky_l\right)\right)\right) =$$

$$= E\left(\frac{1}{2N(N-1)}\left(\sum_{h=1}^{H} 2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left((n_h-1)\sum_{k=1}^{N_h} y_k^2 I_{hk} - \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l I_{hk} I_{hl}\right)\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H} 2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left((n_h-1)\sum_{k=1}^{N_h} y_k^2 EI_{hk} - \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l EI_{hk} I_{hl}\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H} 2\frac{N_h(N_h-1)}{n_h(n_h-1)}\left((n_h-1)\sum_{k=1}^{N_h} y_k^2 \frac{n_h}{N_h} - \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l \frac{n_h(n_h-1)}{N_h(N_h-1)}\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H} 2\left((N_h-1)\sum_{k=1}^{N_h} y_k^2 - \sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l\right)\right) =$$

$$= \frac{1}{2N(N-1)}\sum_{h=1}^{H}\sum_{k,l=1}^{N_h}(y_k-y_l)^2.$$

$$E\left(\frac{1}{2N(N-1)}\left(\sum_{\substack{h,f=1\\h\neq f}}^{H} \frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k\in s_h} y_k^2 - 2\sum_{k\in s_h} y_k \sum_{l\in s_f} y_l + n_h \sum_{l\in s_f} y_l^2\right)\right)\right) =$$

$$= E\left(\frac{1}{2N(N-1)}\left(\sum_{\substack{h,f=1\\h\neq f}}^{H} \frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k=1}^{N_h} y_k^2 I_{hk} - 2\sum_{k=1}^{N_h} y_k I_{hk} \sum_{l=1}^{N_f} y_l I_{fl} + n_h \sum_{l=1}^{N_f} y_l^2 I_{fl}\right)\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{\substack{h,f=1\\h\neq f}}^{H} \frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k=1}^{N_h} y_k^2 EI_{hk} - 2\sum_{k=1}^{N_h} y_k EI_{hk} \sum_{l=1}^{N_f} y_l EI_{fl} + n_h \sum_{l=1}^{N_f} y_l^2 EI_{fl}\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{\substack{h,f=1\\h\neq f}}^{H} \frac{N_h}{n_h}\frac{N_f}{n_f}\left(n_f\sum_{k=1}^{N_h} y_k^2 \frac{n_h}{N_h} - 2\sum_{k=1}^{N_h} y_k \frac{n_h}{N_h} \sum_{l=1}^{N_f} y_l \frac{n_f}{N_f} + n_h \sum_{l=1}^{N_f} y_l^2 \frac{n_f}{N_f}\right)\right) =$$

$$= \frac{1}{2N(N-1)}\left(\sum_{\substack{h,f=1\\h\neq f}}^{H} \left(N_f\sum_{k=1}^{N_h} y_k^2 - 2\sum_{k=1}^{N_h} y_k \sum_{l=1}^{N_f} y_l + N_h \sum_{l=1}^{N_f} y_l^2\right)\right) =$$

$$= \frac{1}{2N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}(y_k-y_l)^2.$$

čia

$$I_{hk} = \begin{cases} 1, & \textbf{kai } k \in s_h \\ 0, & \textbf{kai } k \notin s_h \end{cases}$$

$$k = 1, ..., N_h, \quad h = 1, ..., H.$$

Taigi įvertinio $\widehat{S}_2^2$ vidurkis yra

$$E\widehat{S}_2^2 = \frac{1}{2N(N-1)}\left(\sum_{h=1}^{H}\sum_{k,l=1}^{N_h}(y_k - y_l)^2 + \sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}(y_k - y_l)^2\right) =$$

$$= \frac{1}{2N(N-1)}\sum_{h=1}^{H}\sum_{f=1}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f}(y_k - y_l)^2 =$$

$$= \frac{1}{2N(N-1)}\sum_{k,l=1}^{N}(y_k - y_l)^2 = S_2^2.$$

Iš čia išplaukia, kad įvertinys $\widehat{S}_2^2$ yra nepaslinktasis. □

3. Teiginys. Dispersijos S^2 įvertinys $\widehat{S}_3^2$ yra nepaslinktasis.

Įrodymas :

$$\widehat{S}_4^2 = \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}y_k^2 - \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{\substack{k,l\in s_h\\k\neq l}}y_k y_l -$$

$$-\frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f}y_k y_l.$$

$$E\left(\frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}y_k^2\right) = E\left(\frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2 I_{hk}\right) = \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2 E I_{hk} =$$

$$= \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k=1}^{N_h}y_k^2\frac{n_h}{N_h} = \frac{1}{N}\sum_{h=1}^{H}\sum_{k=1}^{N_h}y_k^2 = \frac{1}{N}\sum_{k=1}^{N}y_k^2.$$

$$E\left(\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{\substack{k,l\in s_h\\k\neq l}} y_k y_l\right) = E\left(\frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l I_{hk} I_{hl}\right) =$$

$$= \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l E I_{hk} I_{hl} =$$

$$= \frac{1}{N(N-1)}\sum_{h=1}^{H}\frac{N_h(N_h-1)}{n_h(n_h-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l \frac{n_h(n_h-1)}{N_h(N_h-1)} = \frac{1}{N(N-1)}\sum_{h=1}^{H}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l.$$

$$E\left(\frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k\in s_h}\sum_{l\in s_f} y_k y_l\right) = E\left(\frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f} y_k y_l I_{hk} I_{fl}\right) =$$

$$= \frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f} y_k y_l E I_{hk} I_{fl} = \frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\frac{N_h}{n_h}\frac{N_f}{n_f}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f} y_k y_l \frac{n_h}{N_h}\frac{n_f}{N_f} =$$

$$= \frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f} y_k y_l.$$

čia

$$I_{hk} = \begin{cases} 1, & \textbf{kai } k \in s_h \\ 0, & \textbf{kai } k \notin s_h \end{cases}$$

$$k = 1, ..., N_h, \quad h = 1, ..., H.$$

Įvertinio $\widehat{S}_3^2$ vidurkis yra

$$E\widehat{S}_3^2 = \frac{1}{N}\sum_{k=1}^{N} y_k^2 - \frac{1}{N(N-1)}\sum_{h=1}^{H}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l - \frac{1}{N(N-1)}\sum_{\substack{h,f=1\\h\neq f}}^{H}\sum_{k=1}^{N_h}\sum_{l=1}^{N_f} y_k y_l =$$

$$= \frac{1}{N}\sum_{k=1}^{N} y_k^2 - \frac{1}{N(N-1)}\sum_{h=1}^{H}\sum_{f=1}^{H}\sum_{\substack{k,l=1\\k\neq l}}^{N_h} y_k y_l = \frac{1}{N}\sum_{k=1}^{N} y_k^2 - \frac{1}{N(N-1)}\sum_{\substack{k,l=1\\k\neq l}}^{N} y_k y_l = S_3^2.$$

Taigi įvertinys $\widehat{S}_3^2$ yra nepaslinktasis. $\square$

4. Teiginys. Sluoksninės imties su paprastąja atsitiktine imtimi sluoksniuose atveju parametro parametro S^2 įvertinio $\widehat{S}^2$ dispersijos $D\widehat{S}^2$ įvertinys $\widehat{D}\widehat{S}^2$ skaičiuojamas pagal formulę (*).

Įrodymas. Sąlygoje nurodytam imties planui

$$\pi_{hk} = \frac{n_h}{N_h}$$

$$\pi_{hkl} = \frac{n_h(n_h-1)}{N_h(N_h-1)}$$

$$h = 1, ..., H.$$

Tada

$$\widehat{D}\widehat{S}^2 = \sum_{h=1}^{H} \sum_{k,l\in s_h} \left(1 - \frac{\pi_{hk}\pi_{hl}}{\pi_{hkl}}\right) \frac{\widehat{u}_{hk}}{\pi_{hk}} \frac{\widehat{u}_{hl}}{\pi_{hl}} + \sum_{\substack{h,f=1\\h\neq f}}^{H} \sum_{k\in s_h} \sum_{l\in s_f} \left(1 - \frac{\pi_{hk}\pi_{fl}}{\pi_{hk}\pi_{fl}}\right) \frac{\widehat{u}_{hk}}{\pi_{hk}} \frac{\widehat{u}_{fl}}{\pi_{fl}} =$$

$$= \sum_{h=1}^{H} \sum_{k,l\in s_h} \left(1 - \frac{\pi_{hk}\pi_{hl}}{\pi_{hkl}}\right) \frac{\widehat{u}_{hk}}{\pi_{hk}} \frac{\widehat{u}_{hl}}{\pi_{hl}} =$$

$$= \sum_{h=1}^{H} \sum_{k=l\in s_h} \left(1 - \frac{\pi_{hk}\pi_{hk}}{\pi_{hk}}\right) \frac{\widehat{u}_{hk}^2}{\pi_{hk}^2} + \sum_{h=1}^{H} \sum_{\substack{k,l\in s_h\\k\neq l}} \left(1 - \frac{\pi_{hk}\pi_{hl}}{\pi_{hkl}}\right) \frac{\widehat{u}_{hk}}{\pi_{hk}} \frac{\widehat{u}_{hl}}{\pi_{hl}} =$$

$$= \sum_{h=1}^{H} \sum_{k\in s_h} \left(\frac{1}{\pi_{hk}^2} - \frac{1}{\pi_{hk}}\right) \widehat{u}_{hk}^2 + \sum_{h=1}^{H} \sum_{\substack{k,l\in s_h\\k\neq l}} \left(\frac{1}{\pi_{hk}\pi_{hl}} - \frac{1}{\pi_{hkl}}\right) \widehat{u}_{hk}\widehat{u}_{hl} =$$

$$= \sum_{h=1}^{H} \sum_{k\in s_h} \left(\frac{1}{\frac{n_h^2}{N_h^2}} - \frac{1}{\frac{n_h}{N_h}}\right) \widehat{u}_{hk}^2 + \sum_{h=1}^{H} \sum_{\substack{k,l\in s_h\\k\neq l}} \left(\frac{1}{\frac{n_h^2}{N_h^2}} - \frac{1}{\frac{n_h(n_h-1)}{N_h(N_h-1)}}\right) \widehat{u}_{hk}\widehat{u}_{hl} =$$

$$= \sum_{h=1}^{H} \sum_{k\in s_h} \left(\frac{N_h^2}{n_h^2} - \frac{N_h}{n_h}\right) \widehat{u}_{hk}^2 + \sum_{h=1}^{H} \sum_{\substack{k,l\in s_h\\k\neq l}} \left(\frac{N_h^2}{n_h^2} - \frac{N_h(N_h-1)}{n_h(n_h-1)}\right) \widehat{u}_{hk}\widehat{u}_{hl} =$$

$$= \sum_{h=1}^{H} \sum_{k\in s_h} \frac{N_h^2 - N_h n_h}{n_h^2} \widehat{u}_{hk}^2 + \sum_{h=1}^{H} \sum_{\substack{k,l\in s_h\\k\neq l}} \frac{N_h^2(n_h-1) - (N_h^2 - N_h)n_h}{n_h^2(n_h-1)} \widehat{u}_{hk}\widehat{u}_{hl} =$$

$$= \sum_{h=1}^{H} \sum_{k \in s_h} \frac{N_h(N_h - n_h)}{n_h^2} \widehat{u}_{hk}^2 + \sum_{h=1}^{H} \sum_{\substack{k,l \in s_h \\ k \neq l}} \frac{N_h(n_h - N_h)}{n_h^2(n_h - 1)} \widehat{u}_{hk} \widehat{u}_{hl} =$$

$$= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h^2(n_h - 1)} \left((n_h - 1) \sum_{k \in s_h} \widehat{u}_{hk}^2 - \sum_{\substack{k,l \in s_h \\ k \neq l}} \widehat{u}_{hk} \widehat{u}_{hl} \right) =$$

$$= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h^2(n_h - 1)} \left(n_h \sum_{k \in s_h} \widehat{u}_{hk}^2 - \sum_{k \in s_h} \widehat{u}_{hk}^2 - \sum_{\substack{k,l \in s_h \\ k \neq l}} \widehat{u}_{hk} \widehat{u}_{hl} \right) =$$

$$= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h^2(n_h - 1)} \left(n_h \sum_{k \in s_h} \widehat{u}_{hk}^2 - \left(\sum_{k \in s_h} \widehat{u}_{hk} \right)^2 \right) =$$

$$= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} \left(\frac{1}{n_h - 1} \sum_{k \in s_h} \widehat{u}_{hk}^2 - \frac{1}{n_h(n_h - 1)} \left(\sum_{k \in s_h} \widehat{u}_{hk} \right)^2 \right) =$$

$$= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} \widehat{s}_h^2,$$

čia

$$\widehat{s}_h^2 = \frac{1}{n_h - 1} \sum_{k \in s_h} (\widehat{u}_{hk} - \widehat{\mu}_{hk})^2$$

$$\widehat{\mu}_{hk} = \frac{1}{n_h} \sum_{k \in s_h} \widehat{u}_{hk}$$

$$\widehat{u}_{hk} = \frac{1}{N - 1} \left(\left(y_{hk} - \frac{1}{N} \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} y_k \right)^2 - \frac{1}{N - 1} \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} \left(y_k - \frac{1}{N} \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in s_h} y_k \right)^2 \right) . \square$$

5 Saviranko rūšys

Nagrinėjame N dydžio baigtinę populiaciją, kurią padaliname į H nepersidengiančių sluoksnių. h - tajame sluoksnyje yra N_h elementų. Todėl $N_1 + N_2 + ... + N_H = N$.

Kiekviename sluoksnyje nepriklausomai nuo kitų sluoksnių, yra išrenkama paprastoji atsitiktinė (grąžintinė arba negrąžintinė) imtis. Imties dydis h -tajame sluoksnyje yra n_h ir todėl visos imties dydis $n = n_1 + n_2 + ... + n_H$.

Turime mus dominančių skaitinių charakteristikų matavimų vektorių

$$\mathbf{y}_{hi} = (y_{1hi}, y_{2hi}, ..., y_{\tau hi})^T,$$

kur h nurodo sluoksnį ir i nurodo elementą h- tajame sluoksnyje.

Mus dominantis populiacijos parametras

$$\theta = \theta(S),$$

kur

$$S = \{\mathbf{y}_{hi} : h = 1, ..., H; i = 1, ..., N_h\}.$$

Mus dominančio populiacijos parametro įvertinys

$$\widehat{\theta} = \widehat{\theta}(s),$$

kur

$$s = \{\mathbf{y}_{hi} : h = 1, ..., H; i = 1, ..., n_h\}.$$

5.1 Tradicinė saviranka

Atsižvelgdami į savybę, kad $\mathbf{y}_{hi}$ yra nepriklausomi, vienodai pasiskirstę imtyje $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$, turime paprastai išplėstą tradicinės savirankos metodą [1] sluoksninėms imtims:

1. Išrenkame paprastąją atsitiktinę grąžintinę imtį $\{\mathbf{y}^*_{hi}\}_{i=1}^{n_h}$ iš pradinės imties $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$ h-tajame sluoknyje nepriklausomai nuo kitų sluoksnių, kad gauti

$$s^* = \{\mathbf{y}^*_{hi} : h = 1, ..., H; i = 1, ..., n_h\}$$

[1] angl. naive bootstrap

Tarkime, kad parametro θ įvertinys $\widehat{\theta}^* = \widehat{\theta}(s^*)$.

2. Pakartojame 1-ą žingsnį nepriklausomai didelį skaičių B kartų ir apskaičiuojame parametro θ įvertinius

$$\widehat{\theta}_1^*, \widehat{\theta}_2^*, ..., \widehat{\theta}_B^*.$$

3. Įvertiname įvertinio $\widehat{\theta}$ dispersiją $D\widehat{\theta}$

$$v_b = D_*\widehat{\theta}^* = E_*(\widehat{\theta}^* - E_*\widehat{\theta}^*)^2$$

arba tradicinės saviranko metodu gautas jos įvertinys

$$\widetilde{v}_b = \widehat{D}_*\widehat{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_i^* - \widehat{\theta}_{(\cdot)}^*)^2, \qquad (1)$$

kur

$$\widehat{\theta}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*. \qquad (2)$$

E_* ir D_* nurodo atitinkamai vidurkį ir dispersiją atsižvelgiant į savirankos kartotinę imtį. Lygybėse (1) ir (2) $E_*\widehat{\theta}^*$ ir $\widehat{\theta}_{(\cdot)}^*$ gali būti pakeisti $\widehat{\theta}$.

Paprasčiausiu atveju, kai $\tau = 1$ ir $\widehat{\theta} = \overline{y}$,
čia

$$\overline{y} = \sum_{h=1}^{H} W_h\overline{y}_h, \quad \overline{y}_h = \sum_{i=1}^{n_h}\frac{y_{hi}}{n_h}, \quad W_h = \frac{N_h}{N}$$

Turime

$$v_b = \sum_{h=1}^{H}\frac{W_h^2}{n_h}\left(\frac{n_h-1}{n_h}\right)s_h^2 \qquad (*)$$

čia

$$s_h^2 = \sum_{i=1}^{n_h}\frac{(y_{hi}-\overline{y}_h)^2}{n_h-1}.$$

Lyginant $(*)$ su standartiniu nepaslinktu dispersijos įvertiniu $\widehat{D}\overline{y}$,kur

$$\widehat{D}(\overline{y}) = \sum_{h=1}^{H} W_h^2\frac{1-f_h}{n_h}s_h^2, \qquad f_h = \frac{n_h}{N_h}$$

matome, kad v_b nėra nepaslinktas $D(\overline{y})$ įvertinys ir svarbiu specialiu atveju, kai n_h yra apribotas, nėra suderintas $D(\overline{y})$ įvertinys, kai $H \to \infty$.

To galima išvengti naudojant pataisos daugiklį tuo atveju, jei $n_h = n_0$ ir $f_h = f$ visiems h ir šiuo atveju $\frac{n_0}{n_0-1}(1-f)v_b$ yra suderintas ir nepaslinktas.

5.2 Mastelio pakeitimo saviranka

Daugelis įprastai naudojamų statistikų gali būti užrašomos $\widehat{\theta} = g(\overline{\mathbf{y}})$, t.y. kaip tam tikrų kintamųjų vidurkių funkcija (taip pat santykį, regresijos koeficientą ir koreliacijos koeficientą galima išreikšti tam tikrų kintamųjų vidurkių funkcijomis). Įvertinio $\widehat{\theta} = g(\overline{\mathbf{y}})$ dispersijos vertinimui RAO ir WU (1988) pasiūlė procedūrą, kurios esmė yra tokia - iš pradinės imties išrenkame naują grąžintinę imtį. Tada kiekvienam naujosios imties elementui taikome mastelio pakeitimą ir šiuos elementus naudojame parametro $\widehat{\theta}$ dispersijos vertinimui. Mastelio pakeitimo daugiklis yra parenkamas taip, kad tiesiniu atveju dispersija, gauta panaudojant kartotinę imtį, atitiktų įprastą dispersijos įvertinimą.

Sluoksninių imčių atveju mastelio pakeitimo savirankos metodas [2] yra toks:

1. Išrenkame paprastą atsitiktinę grąžintinę imtį $\{\mathbf{y}^*_{hi}\}_{i=1}^{m_h}$ $(m_h \geq 1)$ iš pardinės imties $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$ h - tajame sluoksnyje nepriklausomai nuo kitų sluoksnių.

Apskaičiuojame

$$\widetilde{\mathbf{y}}_{hi} = \overline{\mathbf{y}}_h + m_h^{\frac{1}{2}}(n_h-1)^{-\frac{1}{2}}(1-\lambda_h f_h)^{\frac{1}{2}}(\mathbf{y}^*_{hi} - \overline{\mathbf{y}}_h), i = 1, ..., m_h, h = 1, ..., H$$

čia

$$\overline{\mathbf{y}}_h = \sum_{i=1}^{n_h} \frac{\mathbf{y}_{hi}}{n_h}, \quad f_h = \frac{n_h}{N_h}$$

$$\lambda_h = \begin{cases} 1, & \text{jei pradinė imtis yra negrąžintinė sluoksnyje h} \\ 0, & \text{jei pradinė imtis yra grąžintinė sluoksnyje h} \end{cases}$$

$$\overline{\widetilde{\mathbf{y}}}_h = \frac{1}{m_h}\sum_{i=1}^{m_h} \widetilde{\mathbf{y}}_{hi} = \overline{\mathbf{y}}_h + m_h^{\frac{1}{2}}(n_h-1)^{-\frac{1}{2}}(1-\lambda_h f_h)^{\frac{1}{2}}(\overline{\mathbf{y}}^*_h - \overline{\mathbf{y}}_h)$$

čia

$$\overline{\mathbf{y}}^*_h = \sum_{i=1}^{m_h} \frac{\mathbf{y}^*_{hi}}{m_h}$$

[2]angl. rescaling bootstrap

$$\overline{\overline{\mathbf{y}}} = \sum_{h=1}^{H} W_h \overline{\overline{\mathbf{y}}}_h, \qquad W_h = \frac{N_h}{N}$$

Parametro θ įvertinys

$$\widehat{\theta}^* = g(\overline{\overline{\mathbf{y}}}).$$

2. Pakartojame 1-ą žingsnį nepriklausomai didelį skaičių B kartų ir apskaičiuojame parametro θ įvertinius

$$\widehat{\theta}_1^*, \widehat{\theta}_2^*, ..., \widehat{\theta}_B^*.$$

3. Įvertiname įvertinio $\widehat{\theta}$ dispersiją $D\widehat{\theta}$

$$v_r = D_* \widehat{\theta}^* = E_*(\widehat{\theta}^* - E_* \widehat{\theta}^*)^2$$

arba mastelio pakeitimo saviranko metodu gautas jos įvertinys

$$\widetilde{v}_r = \widehat{D}_* \widehat{\theta}^* = \frac{1}{B-1} \sum_{i=1}^{B} (\widehat{\theta}_i^* - \widehat{\theta}_{(\cdot)}^*)^2, \qquad (1)$$

kur

$$\widehat{\theta}_{(\cdot)}^* = \frac{1}{B} \sum_{i=1}^{B} \widehat{\theta}_i^*. \qquad (2)$$

E_* nurodo vidurkį atsižvelgiant į saviranko kartotinę imtį. Lygybėse (1) ir (2) $E_* \widehat{\theta}^*$ ir $\widehat{\theta}_{(\cdot)}^*$ gali būti pakeisti $\widehat{\theta}$.

Papraščiausiu atveju, kai $\tau = 1$ ir $\widehat{\theta} = \overline{y}$, tada $v_r = \widehat{D}\overline{y}$ įprastas nepaslinktas dispersijos įvertinys su bet kokiu m_h;

$$v_b = E_*(\overline{\overline{y}} - \overline{y})^2 = E_* \left(\sum_{h=1}^{H} W_h \overline{\overline{y}}_h - \sum_{h=1}^{H} W_h \overline{y}_h \right)^2 =$$

$$= E_* \left(\sum_{h=1}^{H} W_h (\overline{\overline{y}}_h - \overline{y}_h) \right)^2 = E_* \left(\sum_{h=1}^{H} W_h^2 (\overline{\overline{y}}_h - \overline{y}_h)^2 + \sum_{\substack{h,h'=1 \\ h \neq h'}}^{H} W_h W_{h'} (\overline{\overline{y}}_h - \overline{y}_h)(\overline{\overline{y}}_{h'} - \overline{y}_{h'}) \right) =$$

$$= E_* \Bigg(\sum_{h=1}^{H} W_h^2 (\overline{y}_h + m_h^{\frac{1}{2}} (n_h - 1)^{-\frac{1}{2}} (1 - \lambda_h f_h)^{\frac{1}{2}} (\overline{y}_h^* - \overline{y}_h) - \overline{y}_h)^2 +$$

$$+ \sum_{\substack{h,h'=1 \\ h \neq h'}}^{H} W_h W_{h'} (\overline{y}_h + m_h^{\frac{1}{2}} (n_h - 1)^{-\frac{1}{2}} (1 - \lambda_h f_h)^{\frac{1}{2}} (\overline{y}_h^* - \overline{y}_h) - \overline{y}_h) \cdot$$

$$\cdot(\overline{y}_{h'} + m_{h'}^{\frac{1}{2}}(n_{h'}-1)^{-\frac{1}{2}}(1-\lambda_{h'}f_{h'})^{\frac{1}{2}}(\overline{y}_{h'}^{*}-\overline{y}_{h'})-\overline{y}_{h'})\Bigg) =$$

$$= E_*\Bigg(\sum_{h=1}^{H} W_h^2 (m_h^{\frac{1}{2}}(n_h-1)^{-\frac{1}{2}}(1-\lambda_h f_h)^{\frac{1}{2}}(\overline{y}_h^{*}-\overline{y}_h))^2\Bigg) +$$

$$+ \sum_{\substack{h,h'=1\\ h\neq h'}}^{H} W_h W_{h'} (m_h^{\frac{1}{2}}(n_h-1)^{-\frac{1}{2}}(1-\lambda_h f_h)^{\frac{1}{2}})E_*(\overline{y}_h^{*}-\overline{y}_h)(m_{h'}^{\frac{1}{2}}(n_{h'}-1)^{-\frac{1}{2}}(1-\lambda_{h'}f_{h'})^{\frac{1}{2}})E_*(\overline{y}_{h'}^{*}-\overline{y}_{h'}) =$$

$$= \sum_{h=1}^{H} W_h^2 \frac{m_h}{n_h-1}(1-\lambda_h f_h)E_*(\overline{y}_h^{*}-\overline{y}_h)^2 = \sum_{h=1}^{H} W_h^2 \frac{m_h}{n_h-1}(1-\lambda_h f_h)D_*(\overline{y}_h^{*}) =$$

$$= \sum_{h=1}^{H} W_h^2 \frac{m_h}{n_h-1}(1-\lambda_h f_h)D_*\Bigg(\frac{1}{m_h}\sum_{i=1}^{m_h} y_{hi}^{*}\Bigg) = \sum_{h=1}^{H} W_h^2 \frac{m_h}{n_h-1}(1-\lambda_h f_h)\frac{1}{m_h}\frac{n_h-1}{n_h}s_h^2 =$$

$$\sum_{h=1}^{H} W_h^2 \frac{1-\lambda_h f_h}{n_h} s_h^2,$$

čia $s_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi}-\overline{y}_h)^2$.

Su šiuo metodu susijęs klausimas yra m_h parinkimas. RAO ir WU (1988) pasiūlė tokią procedūrą:

Tariame, kad $n_h > 2$. Tada m_h yra gaunamas sulyginus savirankos trečią momentą $E_*(\overline{\widetilde{\mathbf{y}}} - \overline{\mathbf{y}})^3$ su $\overline{\mathbf{y}}$ trečio momento įvertiniu $\widehat{\mu}_3(\overline{\mathbf{y}})$

$$\widehat{\mu}_3(\overline{\mathbf{y}}) = \sum_{h=1}^{H} W_h^3 \frac{(1-\lambda_h f_h)(1-2\lambda_h f_h)}{(n_h-1)(n_h-2)} m_{3h}$$

$$E_*(\overline{\widetilde{\mathbf{y}}} - \overline{\mathbf{y}})^3 = \sum_{h=1}^{H} W_h^3 \frac{(1-\lambda_h f_h)^{\frac{3}{2}}}{m_h^{\frac{1}{2}}(n_h-1)^{\frac{3}{2}}} m_{3h}$$

čia

$$m_{3h} = \sum_{i=1}^{n_h} \frac{(\mathbf{y}_{hi} - \overline{\mathbf{y}}_h)^3}{n_h}$$

yra imties trečias momentas sluoksnyje h.

Tada

$$m_h = \frac{(1-\lambda_h f_h)(n_h-2)^2}{(1-2\lambda_h f_h)^2(n_h-1)} \quad (n_h > 2).$$

Galima pastebėti, kad $m_h \simeq n_h - 3$, kai $n_h \geq 5$.

Jei $\lambda_h = 0$ ir $m_h = n_h - 1$ visiems h, tada mastelio pakeitimo savirankos metodas sutampa su grąžintinės savirankos metodu.

Netinkamas m_h parinkimas gali privesti prie neigiamų $\widehat{\theta}$ reikšmių netgi tuo atveju, kai θ yra teigiamas.

5.3 Veidrodinio sutapimo saviranka

Sitter (1992) pasiūlė savirankos metodą, kurį mes vadiname veidrodinio sutapimo saviranka [3]. Šio metodo esmė - išrenkame negrąžintinį poimtį iš pradinės imties taip, kad atspindėtų pradinio ėmimo schemą. Tada pakeičiame šį poimtį ir pakartojame tai apibrėžtą skaičių kartų k_h. Ši savirankos procedūra kartojama didelį skaičių kartų. k_h reikšmė parenkama taip, kad tiesiniu atveju savirankos metodu gauta dispersija atitiktų įprastą dispersijos įvertinimą.

Sluoksninių imčių atveju veidrodinio sutapimo savirankos metodas yra toks:

Tarkime, kad $N_h = n_h k_h (\leftrightarrow k_h = f_h^{-1})$ ir $n'_h = f_h n_h$, čia $k_h \geq 1$ ir $n'_h \geq 1$ sveiki skaičiai kiekvienam $h = 1, ..., H$. (Jei k_h ir n'_h nėra sveiki skaičiai, tai naudosime randomizaciją, kuri bus pateikta vėliau).

1. Pasirenkame $1 \leq n'_h < n_h$ ir išrenkame paprastąją atsitiktinę negrąžintinę imtį dydžio n'_h iš $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$, kad gauti $\mathbf{y}^*_{h1}, \mathbf{y}^*_{h2}, ..., \mathbf{y}^*_{hn'_h}$.

2. Pakartojame 1-ą žingsnį nepriklausomai k_h kartų

$$k_h = \frac{n_h(1 - f_h^*)}{n'_h(1 - \lambda_h f_h)}$$

$$\lambda_h = \begin{cases} 1, & \text{jei pradinė imtis yra negrąžintinė sluoksnyje h} \\ 0, & \text{jei pradinė imtis yra grąžintinė sluoksnyje h} \end{cases}$$

pakeisdami n'_h dydžio imtį kiekvieną kartą (t.y. grąžindami ankstesnio ėmimo elementus į pradinę imtį ir rinkdami imtį iš naujo), kad gauti $\mathbf{y}^*_{h1}, \mathbf{y}^*_{h2}, ..., \mathbf{y}^*_{hn^*_h}$, kur $f^*_h = \frac{n'_h}{n_h}$ ir $n^*_h = k_h n'_h$.

3. Pakartojame 1-ą ir 2-ą žingsnį nepriklausomai kiekvienam sluoksniui, kad gauti

$$s^* = \{\mathbf{y}^*_{hi} : h = 1, ..., H; i = 1, ..., n^*_h\}.$$

Tegul parametro θ įvertinys $\widehat{\theta}^* = \widehat{\theta}(s^*)$.

[3]angl. mirror-match bootstrap

4. Pakartojame 1-3 žingsnį nepriklausomai didelį skaičių B kartų ir apskaičiuojame parametro θ įvertinius

$$\widehat{\theta}_1^*, \widehat{\theta}_2^*, ..., \widehat{\theta}_B^*.$$

5. Įvertiname įvertinio $\widehat{\theta}$ dispersiją $D\widehat{\theta}$

$$v_m = D_*\widehat{\theta}^* = E_*(\widehat{\theta}^* - E_*\widehat{\theta}^*)^2$$

arba veidrodinio sutapimo savirankos metodu gautas jos įvertinys

$$\widetilde{v}_m = \widehat{D}_*\widehat{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_i^* - \widehat{\theta}_{(\cdot)}^*)^2, \qquad (1)$$

kur

$$\widehat{\theta}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*. \qquad (2)$$

E_* nurodo vidurkį atsižvelgiant į savirankos kartotinę imtį. Lygybėse (1) ir (2) $E_*\widehat{\theta}^*$ ir $\widehat{\theta}_{(\cdot)}^*$ gali būti pakeisti $\widehat{\theta}$.

Paprasčiausiu atveju, kai $\tau = 1$ ir $\widehat{\theta} = \overline{y}$, tada $v_m = \widehat{D}\overline{y}$ įprastas nepaslinktas dispersijos įvertinys. Jei $f_h \geq \frac{1}{n_h}$, tada pasirinkimas $n_h' = f_h n_h$ reiškia, kad poimčio ėmimo lygis $f_h^* = f_h$ sutampa su pradinės imties ėmimo lygiu nagrinėjamame sluoksnyje.

Su šiuo metodu susijęs klausimas yra n_h' parinkimas. Todėl n_h' parenkame taip , kad savirankos trečias momentas, kai $\tau = 1$, $E_*(\overline{y}^* - \overline{y})^3$ yra lygus $\widehat{\mu}_3(\overline{y})$.

$$\overline{y} = \sum_{h=1}^{H} W_h \overline{y}_h = \sum_{h=1}^{H} W_h \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$$

$$\overline{y}^* = \sum_{h=1}^{H} W_h \overline{y}_h^* = \sum_{h=1}^{H} W_h \sum_{j=1}^{k_h} \sum_{l=1}^{n_h'} \frac{y_{hjl}^*}{k_h n_h'}$$

Tardami, kad k_h ir n_h' yra sveiki skaičiai, gausime, kad bendras imties dydis h-tajame sluoksnyje bus

$$n_h^* = k_h n_h' = \frac{n_h(1 - f_h^*)}{(1 - \lambda_h f_h)}$$

kur $f_h^* = \frac{n_h'}{n_h}$.

Galime pastebėti, kad jei $n'_h = 1, \forall h$, tai šis metodas susiveda į grąžintinės savirankos metodą.

Jei taikydami aprašytą metodą sluoksninėms imtims pasirenkame n'_h sveiką skaičių tokį, kad

$$1 \leq n'_h \leq \frac{n_h}{2 - \lambda_h f_h},$$

tada

$$k_h = \frac{n_h(1 - f_h^*)}{n'_h(1 - \lambda_h f_h)} \geq 1$$

kiekviename sluoksnyje h.

Jei k_h nėra sveikas skaičius, tegu K_h yra atsitiktinis kintamasis toks, kad

$$P(K_h = k_{h1}) = \frac{(\frac{1}{k_h} - \frac{1}{k_{h2}})}{(\frac{1}{k_{h1}} - \frac{1}{k_{h2}})} = p_h,$$

ir

$$P(K_h = k_{h2}) = 1 - p_h,$$

kur $1 \leq k_{h1} < k_h < k_{h2} < n_h$ ir k_{h1} ir k_{h2} yra sveiki skaičiai ($k_{h1} = \lfloor k_h \rfloor$ ir $k_{h2} = \lceil k_h \rceil$, čia $\lceil \cdot \rceil$ ir $\lfloor \cdot \rfloor$ reiškia atitinkamai didžiausias sveikas skaičius mažesnis nei, ir mažiausias sveikas skaičius didesnis nei).

Toliau taikome veidrodinio sutapimo savirankos metodo procedūrą naudodami K_h vietoj k_h. Randomizacija privalo būti taikoma nepriklausomai kiekviename sluoksnyje ir kartojama kiekviename savirankos ėmime.

Jei aprašytame savirankos metode naudojame pasiūlytą n'_h reikšmę $n'_h = f_h n_h$, tada yra įmanomas variantas, kad abu k_h ir n'_h yra ne sveiki skaičiai. Šiuo atveju yra daug būdų atlikti randomizaciją taip, kad $D_* \bar{y}^* = D\bar{y}$. Kadangi k_h randomizacija jau aprašyta ankščiau bet kokiam sveikam skaičiui n'_h, todėl dabar aprašysime n'_h randomizaciją.

Papraščiausias metodas būtų leisti $n'_{h1} = \lfloor n'_h \rfloor$ ir $n'_{h2} = \lceil n'_h \rceil$ ir paimti n'_{h1} su tikimybe p ir n'_{h2} su tikimybe $1 - p$, kur

$$p = n'_{h2} - n'_h$$

čia $\lceil \cdot \rceil$ ir $\lfloor \cdot \rfloor$ reiškia atitinkamai didžiausias sveikas skaičius mažesnis nei, ir mažiausias sveikas skaičius didesnis nei. Tai atliekame nepriklausomai kiekviename sluoksnyje.

Duodami n'_h reikšmę gautą atlikus randomizaciją, galime atlikti k_h randomizaciją, kuri buvo aprašyta ankščiau fiksuotam n'_h. Šis metodas negali būti taikomas tik tuo atveju, jei $\frac{\lceil n'_h \rceil}{n_h} > \frac{1}{2}$, nes tada k_h reikšmė gaunama mažesnė už vienetą.

5.4 Negrąžintinė saviranka

Pirmiausiai tariame, kad pradinė imtis yra negrąžintinė t.y. $\lambda_h = 1$ visiems h.

$$\lambda_h = \begin{cases} 1, & \text{jei pradinė imtis yra negrąžintinė sluoksnyje h} \\ 0, & \text{jei pradinė imtis yra grąžintinė sluoksnyje h} \end{cases}$$

Negrąžintinės savirankos metodą [4] pasiūlė Gross (1980) ir Chao ir Lo (1985) paprastosios atsitiktinės imties atveju, o Bickel ir Freedman (1984) išplėtojo šį metodą sluoksninėms imtims.

Šio metodo esmė - tariame, kad $N_h = k_h n_h$, kur k_h yra kažkoks sveikas skaičius. Negrąžintinės savirankos metode sukuriama N_h dydžio pseudopopuliacija kartojant pradinę imtį sluoksnyje h tiksliai k_h kartų. Tada išrenkama n_h dydžio paprastoji atsitiktinė negrąžintinė imtis iš pseudopopuliacijos nepriklausomai kiekviename sluoksnyje. Parametro θ dispersijos įvertinimas apskaičiuojamas pagal tas pačias formules kaip ir paprasto savirankos metodo atveju. Nors šis metodas atrodo tinkamas, tačiau jis gali neduoti suderinto dispersijos įvertinio. Bickel ir Freedman (1984) nagrinėjo randomizaciją tarp dviejų pseudopopuliacijų kiekviename sluoksnyje, kuri padėtų pataisyti šią problemą, tačiau jų metodas nėra visada taikomas.

Sitter (1992) pasiūlė tokį modifikuotą negrąžintinės savirankos metodą sluoksninėms imtims:

Tegul $n'_h = n_h - (1 - f_h)$ ir $k_h = \frac{N_h}{n_h}\left(1 - \frac{1-f_h}{n_h}\right)$, $h = 1, ..., H$, $f_h = \frac{n_h}{N_h}$. (#)

1. Sluoksnyje h sukuriame pseudopopuliaciją kartodami imtį $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$ k_h kartų. Tai atliekame kiekviename sluoksnyje.

2. Išrenkame paprastąją atsitiktinę negrąžintinę imtį dydžio n'_h iš pseudopopuliacijos sluoksnio h, kad gauti $\{\mathbf{y}^*_{hi}\}_{i=1}^{n'_h}$, $h = 1, ..., H$.

Tegul parametro θ įvertinys $\widehat{\theta}^* = \widehat{\theta}(s^*)$, kur

$$s^* = \{\mathbf{y}^*_{hi} : h = 1, ..., H; i = 1, ..., n'_h\}.$$

3. Pakartojame 1-ą ir 2-ą žingsnį nepriklausomai didelį skaičių B kartų ir apskaičiuojame parametro θ įvertinius

$$\widehat{\theta}^*_1, \widehat{\theta}^*_2, ..., \widehat{\theta}^*_B.$$

[4] angl. without-replacement bootstrap (BWO)

4. Įvertiname įvertinio $\widehat{\theta}$ dispersiją $D\widehat{\theta}$

$$v_{BWO} = D_*\widehat{\theta}^* = E_*(\widehat{\theta}^* - E_*\widehat{\theta}^*)^2$$

arba negrąžintinės savirankos metodu gautas jos įvertinys

$$\widetilde{v}_{BWO} = \widehat{D}_*\widehat{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_i^* - \widehat{\theta}_{(\cdot)}^*)^2, \qquad (1)$$

kur

$$\widehat{\theta}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*. \qquad (2)$$

E_* nurodo vidurkį atsižvelgiant į savirankos kartotinę imtį. Lygybėse (1) ir (2) $E_*\widehat{\theta}^*$ ir $\widehat{\theta}_{(\cdot)}^*$ gali būti pakeisti $\widehat{\theta}$.

Paprasčiausiu atveju, kai $\tau = 1$ ir $\widehat{\theta} = \overline{y}$, tada $v_{BWO} = \widehat{D}\overline{y}$ įprastas dispersijos įvertinys.

Šiame metode n_h' ir k_h yra parinkti taip, kad patenkintų lygtis

$$f_h^* = f_h \quad ir \quad D_*(\overline{\mathbf{y}}_h^*) = \frac{(1-f_h)}{n_h}s_h^2$$

kur $f_h^* = \frac{n_h'}{k_h n_h}$ yra imties dalis sluoksnyje h.

Visa tai turi prasmės, jei k_h ir n_h'abu yra sveiki skaičiai $\forall h$ ir, žinoma, $0 \leq f_h \leq 1, n_h' = n_h - (1-f_h)$ nėra sveiki skaičiai tik tada, kai $f_h = 0$ arba 1.

Kad to išvengti, naudosime randomizaciją. Kadangi ta pati procedūra yra taikoma nepriklausomai kiekviename sluoksnyje, tai aprašysime procedūrą vienam sluoksniui:

Tegul $k_1 = \lceil k \rceil$ ir $k_2 = \lfloor k \rfloor$ (čia $\lceil\cdot\rceil$ ir $\lfloor\cdot\rfloor$ reiškia atitinkamai didžiausias sveikas skaičius mažesnis nei, ir mažiausias sveikas skaičius didesnis nei), kur k yra duotas (#) ir tegul $n_1' = n-1$ ir $n_2' = n$. Taip pat tegul

$$p = \frac{(1-f)/n(n-1) - a_2}{a_1 - a_2}$$

kur $a_j = \{k_j(1 - n_j'/nk_j)\}/\{n_j'(nk_j - 1)\}, j = 1, 2$. Tada šiame sluoksnyje sudaroma pseudopopuliacija kartojant imtį k_1 kartų ir antrame žingsnyje renkama paprastoji atsitiktinė negrąžintinė n_1' didumo imtis, arba pseudopopuliacija sudaroma pradinę imtį pakartojant k_2 kartų ir antrame žinsnyje renkama dydžio n_2' savirankos imtis.

Variantas (k_1, n'_1) pasirenkamas atsitiktinai su tikimybe p, o variantas (k_2, n'_2) su tikimybe $1-p$.

Toliau 3 ir 4 žingsniai atliekami taip pat.

Taigi toks modifikuotas negrąžintinės savirankos metodas duoda suderintus savirankos įvertinius.

5.5 Grąžintinė saviranka

Grąžintinės savirankos metodas [5] savo esme yra artimas paprastajam savirankos metodui. Pagrindinis skirtumas yra tas, kad išrenkama paprastoji atsitiktinė grąžintinė imtis dydžio m_h, o ne n_h iš pradinės imties kiekviename sluoksnyje.

McCarthy ir Snowden (1985) pasiūlė tokį grąžintinės savirankos metodą sluoksninių imčių atveju:

1. Išrenkame paprastą atsitiktinę grąžintinę imtį$\{\mathbf{y}^*_{hi}\}_{i=1}^{m_h}$ iš $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$, $h = 1, ..., H$ nepriklausomai kiekviename sluoknyje, kur

$$m_h = \frac{n_h - 1}{1 - \lambda_h f_h}$$

$$\lambda_h = \begin{cases} 1, & \text{jei pradinė imtis yra negrąžintinė sluoksnyje h} \\ 0, & \text{jei pradinė imtis yra grąžintinė sluoksnyje h} \end{cases}$$

$$f_h = \frac{n_h}{N_h},$$

kad gauti

$$s^* = \{\mathbf{y}^*_{hi} : h = 1, ..., H; i = 1, ..., m_h\}$$

Tarkime, kad parametro θ įvertinys $\widehat{\theta}^* = \widehat{\theta}(s^*)$.

2. Pakartojame 1-ą žingsnį nepriklausomai didelį skaičių B kartų ir apskaičiuojame parametro θ įvertinius

$$\widehat{\theta}^*_1, \widehat{\theta}^*_2, ..., \widehat{\theta}^*_B.$$

3. Įvertiname įvertinio $\widehat{\theta}$ dispersiją $D\widehat{\theta}$

$$v_{BWR} = D_*\widehat{\theta}^* = E_*(\widehat{\theta}^* - E_*\widehat{\theta}^*)^2$$

[5] angl. with-replacement bootstrap (BWR)

arba grąžintinės savirankos metodu gautas jos įvertinys

$$\widetilde{v}_{BWR} = \widehat{D}_*\widehat{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_i^* - \widehat{\theta}_{(\cdot)}^*)^2, \qquad (1)$$

kur

$$\widehat{\theta}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*. \qquad (2)$$

E_* nurodo vidurkį atsižvelgiant į savirankos kartotinę imtį. Lygybėse (1) ir (2) $E_*\widehat{\theta}^*$ ir $\widehat{\theta}_{(\cdot)}^*$ gali būti pakeisti $\widehat{\theta}$.

Jei pradinė imtis yra grąžintinė, tada $m_h = n_h - 1$, šiuo atveju tą patį buvo pasiūlęs Efron (1982). Tuo atveju, kai m_h nėra sveikas skaičius, naudojame randomizaciją, kuri buvo aprašyta anksčiau (analogiškai kaip k_h iš skyriaus 5.3).

6 Eksperimentinė dalis

Šiame darbe yra nagrinėjamos šešios populiacijos. Pirmų dviejų populiacijų duomenys yra sugeneruoti pagal normalųjį dėsnį, kurio pasiskirstymo funkcija yra

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} \,\mathrm{d}t,$$

kur $a = 0$ ir $\sigma = 1$. Po to, duomenys paslinkti taip, kad visi elementai būtų teigiami.

Sekančių dviejų populiacijų duomenys yra sugeneruoti pagal eksponentinį dėsnį, kurio pasiskirstymo funkcija yra

$$F(x) = 1 - e^{-\alpha x}, \quad x \geq 0$$

kur $\alpha = 1$.

Likusių dviejų populiacijų duomenys yra dviejų ekonominių veiklų įmonių pajamos per metus, išreikštos šimtais tūkstančių litų (duomenys pateikti priede).

Taigi galime sudaryti tokią lentelę apie populiacijas:

Populiacija	Duomenys	Elementų skaičius N
I	Generuoti pagal normalųjį dėsnį	200
II	Generuoti pagal normalųjį dėsnį	300
III	Generuoti pagal eksponentinį dėsnį	200
IV	Generuoti pagal eksponentinį dėsnį	300
V	Realūs	151
VI	Realūs	132

Kadangi žinome visas populiacijos kintamojo reikšmes, todėl galime paskaičiuoti tikrąsias kintamojo dispersijos reikšmes :

Populiacija	Dispersija (S^2)
I	0,971221
II	0,940768
III	0,909388
IV	0,910379
V	19,54684
VI	14,08336

Žinodami tikrąsias populiacijos kintamojo dispersijos reikšmes, mes vertinsime kintamojo dispersiją ir dispersijos įvertinio dispersiją naudodami savirankos metodus, kurie aprašyti penktame skyriuje.

Taikant kiekvieną savirankos metodą, iš kiekvienos populiacijos yra renkama sluoksninė imtis su paprastąja atsitiktine imtimi sluoksniuose. Imties dydžiai nurodyti šioje lentelėje:

Populiacija	I	II	III	IV	V	VI
Imties dydis n	30	30	30	30	30	30
	40	40	40	40	50	50
	50	50	50	50	80	80
	80	100	80	100		

Kiekviename sluoksnyje imties dydis nustatomas remiantis Neimano optimalaus imties paskirstymo principu:

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h},$$

čia N_h- elementų skaičius sluoksnyje h, $h = 1, ..., H$, H- sluoksnių skaičius.

Imties dydžiai populiacijų sluoksniuose yra pateikti lentelėje:

Populiacija	Sluoksnis	Elementų skaičius sluoksnyje	Imties dydis sluoksnyje, kai				
			$n = 30$	$n = 40$	$n = 50$	$n = 80$	$n = 100$
I	1	85	18	24	29	47	
	2	63	4	5	7	10	
	3	52	8	11	14	23	
II	1	101	14	20	24		49
	2	77	5	6	8		15
	3	66	5	6	8		15
	4	56	6	8	10		21
III	1	139	20	26	33	52	
	2	40	5	7	8	13	
	3	21	5	7	9	15	
IV	1	181	16	23	29		57
	2	58	4	5	6		12
	3	38	3	3	4		9
	4	23	7	9	11		22
VI	1	118	19		32	51	
	2	21	7		11	18	
	3	12	4		7	11	
VII	1	101	21		35	59	
	2	20	4		7	11	
	3	11	5		8	10	

Dispersijai S^2 vertinti taikomas toks dispersijos įvertinys $\widehat{S}^2$:

$$\widehat{S}^2 = \frac{1}{N-1}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h}\left(y_k - \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} y_k\right)^2$$

ir dispersijos įvertinio $\widehat{S}^2$ dispersijos $D\widehat{S}^2$ įvertinys $\widehat{D}\widehat{S}^2$:

Savirankos rūšys	Dispersijos įvertinio formulė
Mastelio pakeitimo saviranka	$\widehat{D}\widehat{S}^2 = \frac{1}{B-1}\sum_{i=1}^{B}(\widehat{S}_i^{2*} - \widehat{S}_{(\cdot)}^{2*})^2$ $\widehat{S}_{(\cdot)}^{2*} = \frac{1}{B}\sum_{i=1}^{B}\widehat{S}_i^{2*}$
Tradicinė saviranka, veidrodinio sutapimo saviranka, negrąžintinė saviranka, grąžintinė saviranka	$\widehat{D}\widehat{S}^2 = \frac{1}{B}\sum_{i=1}^{B}(\widehat{S}_i^{2*} - \widehat{S}_{(\cdot)}^{2*})^2$ $\widehat{S}_{(\cdot)}^{2*} = \frac{1}{B}\sum_{i=1}^{B}\widehat{S}_i^{2*}$

Šiame darbe yra paskaičiuojamas ir populiacijos dispersijos variacijos koeficientas, kuris yra gana universalus įvertinio kokybės matas nepriklausantis nuo matavimo vienetų. Kuo šis koeficientas arba jo įvertis mažesnis, tuo tikslesnis įvertinys arba konkretus įvertis. Laikoma, kad įvertis yra pakankamai tikslus, jei jo variacijos koeficiento įvertis neviršyja 0,03.

Populiacijos dispersijos įvertinio variacijos koeficiento įvertinys yra skaičiuojamas pagal formulę:

$$\widehat{cv}(\widehat{S}^2) = \frac{\sqrt{\widehat{D}\widehat{S}^2}}{\widehat{S}^2}.$$

Tam, kad pažiūrėti ar savirankos metodai duoda pakankamai tikslią dispersijos įvertinio dispersiją, šiame darbe dar apskaičiuojame tikrąją dispersijos įvertinio dispersijos reikšmę pagal formulę (*) iš ketvirto skyriaus.

Sekančiame skyriuje yra pateikti eksperimentinės dalies rezultatai.

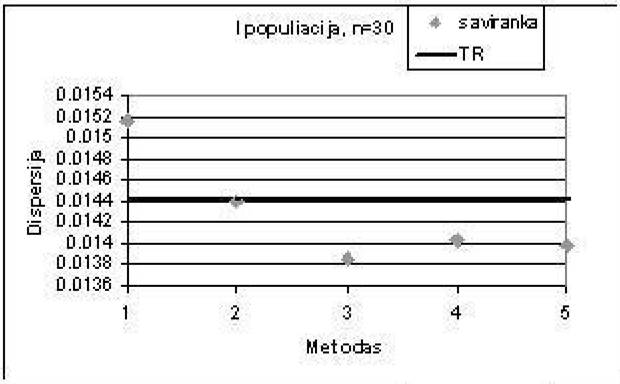
7 Rezultatai

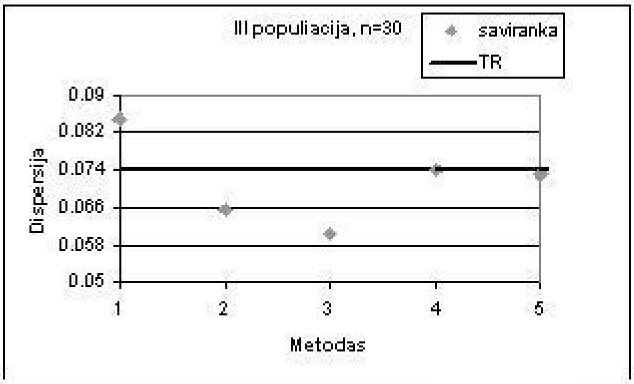
Visi eksperimentinės dalies skaičiavimai įskaitant ir duomenų generavimą atlikti programos SAS 9.1 pagalba.

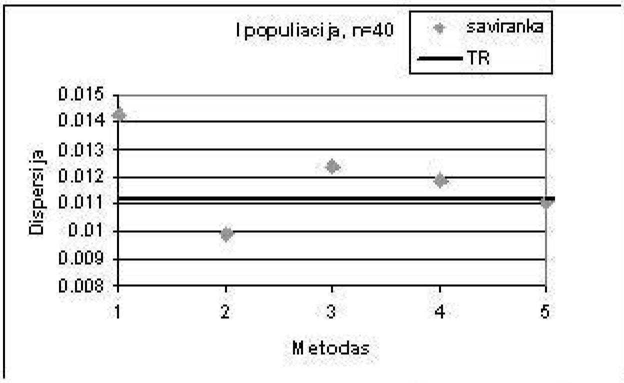
Pirmiausiai apskaičiavome dispersijos įvertinio reikšmes kiekvienai populiacijai ir kiekvienai imčiai, rezultatai pateikti lentelėje:

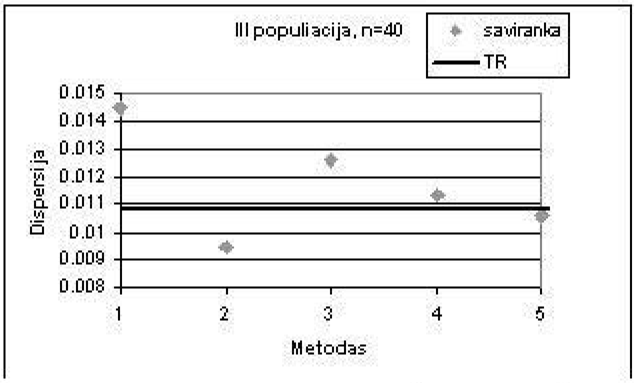
Populiacija	Imties dydis	Dipersijos įvertinys ($\widehat{S}^2$)
I	$n = 30$	0,67534
	$n = 40$	0,73574
	$n = 50$	0,86777
	$n = 80$	0,82802
II	$n = 30$	0,78177
	$n = 40$	0,76597
	$n = 50$	0,90535
	$n = 100$	0,90466
III	$n = 30$	1,06320
	$n = 40$	0,80659
	$n = 50$	0,789298
	$n = 80$	0,92255
IV	$n = 30$	0,82919
	$n = 40$	1,01577
	$n = 50$	1,15267
	$n = 100$	0,86175
IV	$n = 30$	14,7306
	$n = 50$	14,2835
	$n = 80$	14,56612
VI	$n = 30$	13,9205
	$n = 50$	13,4083
	$n = 80$	13,4512

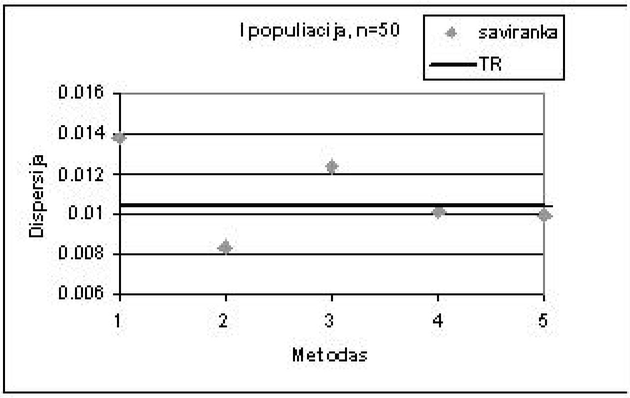
Toliau skaičiavome dispersijos įvertinio dispersijas taikydami savirankos metodus, kai $B = 1000$. Prie tokios B reikšmės įvertinys $\widehat{D}\widehat{S}^2$ jau yra pakankamai stabilus ir didesnės B reikšmės paėmimas nebeturi esminės įtakos. Grafikuose matome kokius rezultatus duoda kiekvienas saviranko metodas ir kiek jų reikšmės skiriasi nuo tikrosios dispersijos įvertinio dispersijos įvertinio reikšmės.

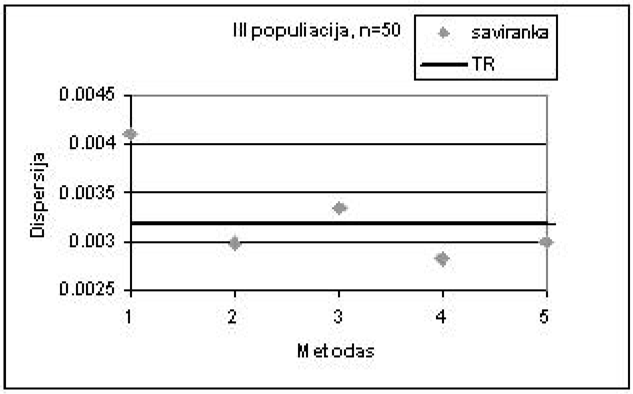

I populiacija, n=30
saviranka
TR
Dispersija
0.0154
0.0152
0.015
0.0148
0.0146
0.0144
0.0142
0.014
0.0138
0.0136
1 2 3 4 5
Metodas

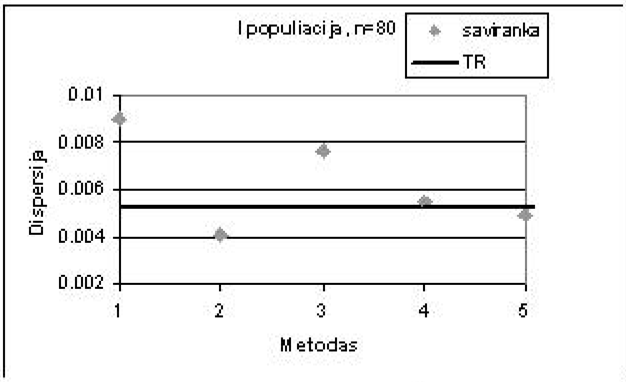

III populiacija, n=30
saviranka
TR
Dispersija
0.09
0.082
0.074
0.066
0.058
0.05
1 2 3 4 5
Metodas

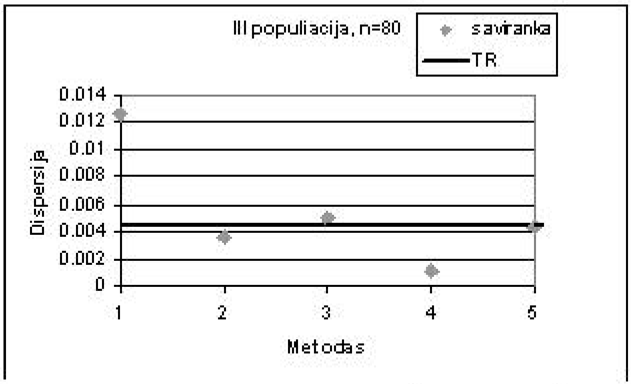

I populiacija, n=40
saviranka
TR
Dispersija
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.008
1 2 3 4 5
Metodas

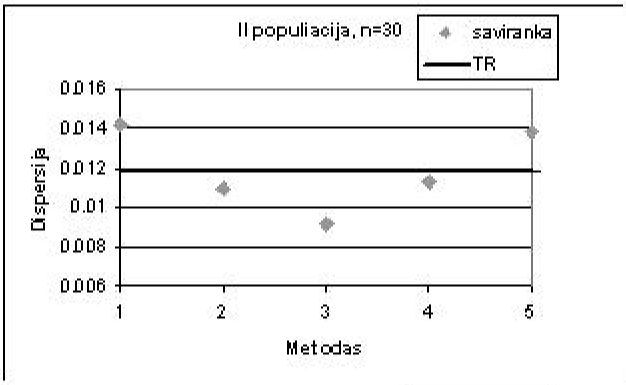

III populiacija, n=40
saviranka
TR
Dispersija
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.008
1 2 3 4 5
Metodas

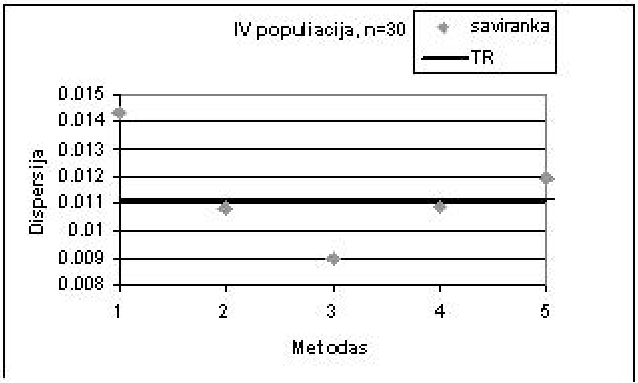

I populiacija, n=50
saviranka
TR
Dispersija
0.016
0.014
0.012
0.01
0.008
0.006
1 2 3 4 5
Metodas

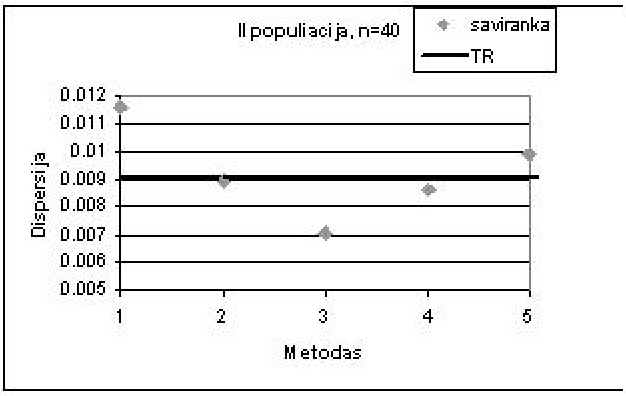

III populiacija, n=50
saviranka
TR
Dispersija
0.0045
0.004
0.0035
0.003
0.0025
1 2 3 4 5
Metodas

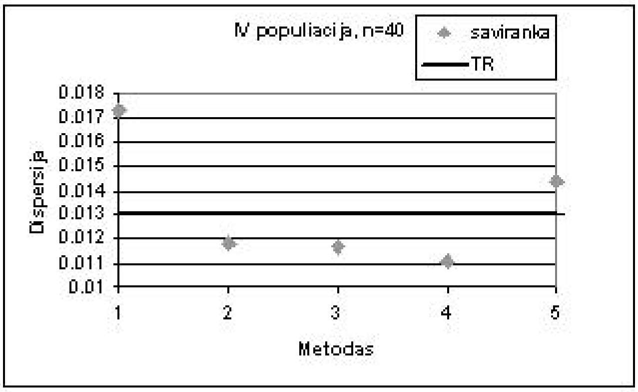

I populiacija, n=80
saviranka
TR
Dispersija
Metodas

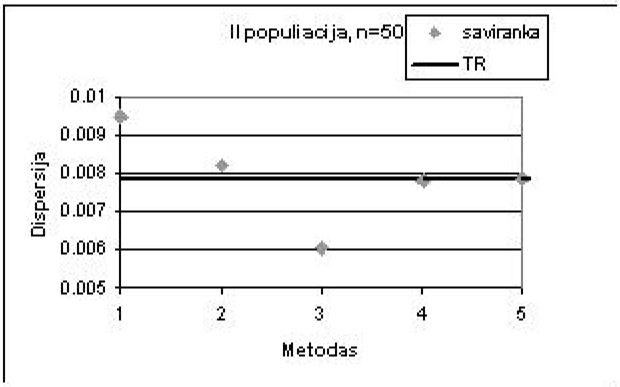

III populiacija, n=80
saviranka
TR
Dispersija
Metodas

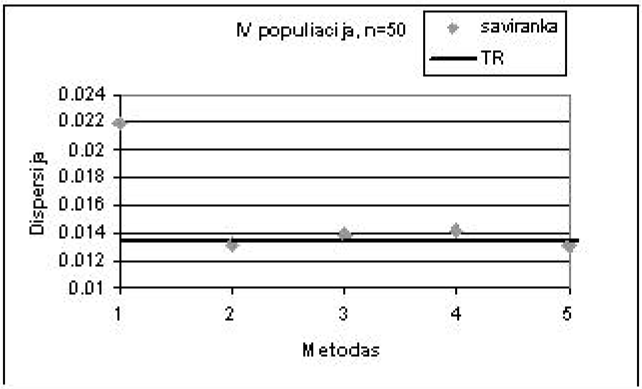

II populiacija, n=30
saviranka
TR
Dispersija
Metodas

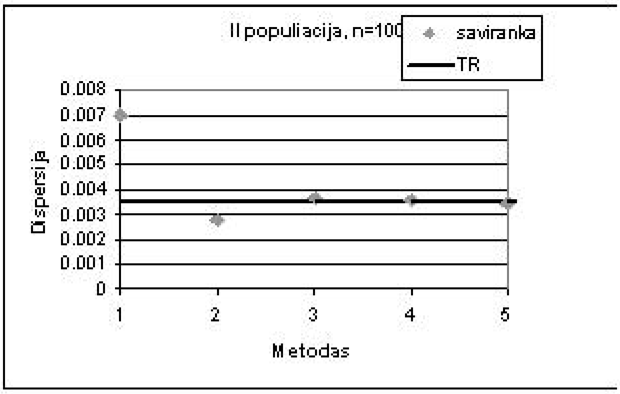

IV populiacija, n=30
saviranka
TR
Dispersija
Metodas

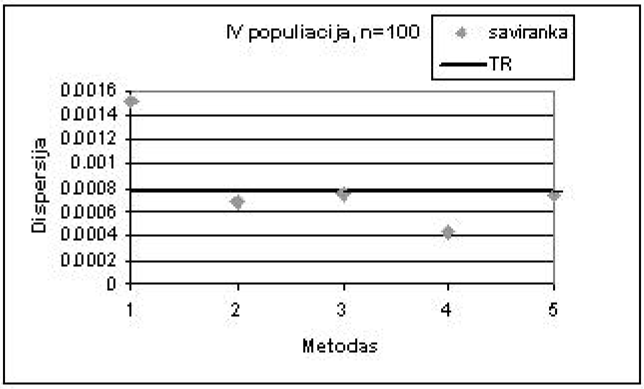

II populiacija, n=40
saviranka
TR
Dispersija
Metodas


IV populiacija, n=40
saviranka
TR
Dispersija
Metodas









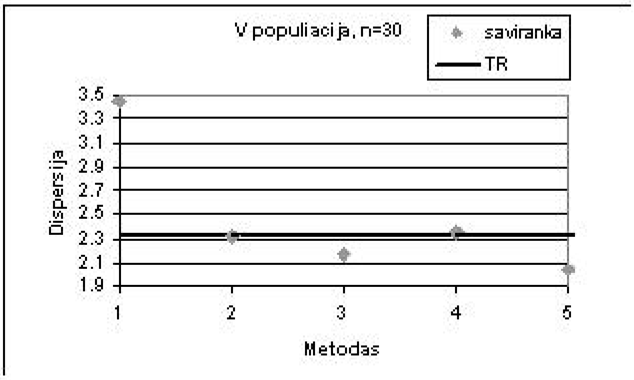
Iš pateiktų grafikų matome, kad esant tokiam pačiam populiacijos dydžiui ir tiems patiems imčių dydžiams, bet skiriantis duomenų pasiskirstymui, savirankos metodai vertina panašiai ta prasme, kad išlaiko metodų išsidėstymą t.y. tie patys metodai duoda didesnę ar mažesnę reikšmę. Aišku, atskiroms imtims ši tendencija gali truputį pasikeisti. Jei lyginsime savirankos metodus su tikrąja reikšme, rezultatai jau išsiskirs daugiau.

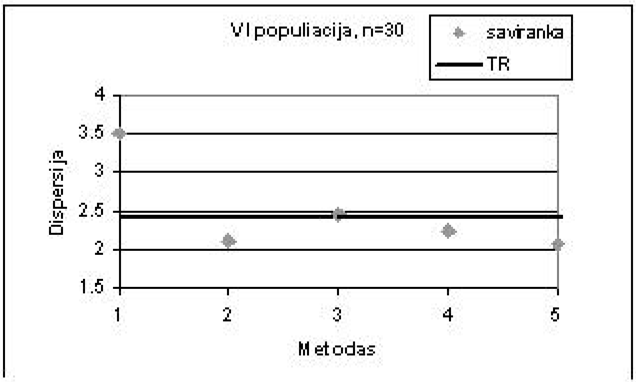
Visoms populiacijoms pirmas savirankos metodas (tradicinė saviranka) duoda ryškiai blogesnį rezultatą, bet tai galima paaiškinti tuo, kad šis metodas remiasi prielaida , kad imtis yra grąžintinė ir nėra jokio pataisos daugiklio renkant negrąžintinę imtį, kurią mes šiame darbe ir renkame. I ir II populiacijoje ketvirtas (negrąžintinė saviranka) ir penktas (grąžintinė saviranka) savirankos metodas duoda geriausius rezultatus lyginant su tikrąja reikšme, o III ir IV populiacijoje - antras (mastelio pakeitimo saviranka) ir penktas (grąžintinė saviranka) savirankos metodas. Žinoma, tai gali keisti atskiroms imtims, bet daugeliu atveju ši tendencija turėtų išsilaikyti. Be to, iš grafikų matome, kad imant skirtingo dydžio, bet vienodai sugeneruotas populiacijas savirankos metodų išsidėstymas nesikeičia.

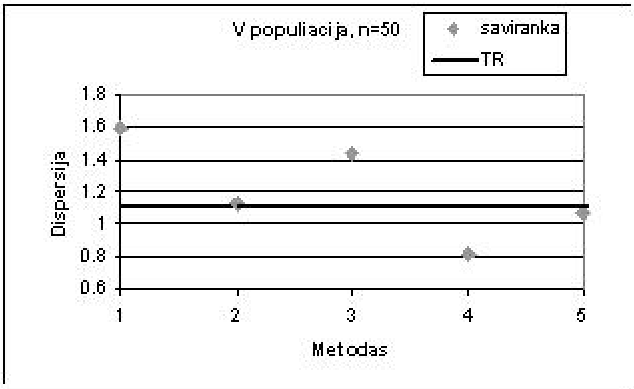
Pažiūrėję į variacijos koeficiento reikšmes kiekvienai populiacijai ir kiekvienai imčiai, rezultatai pateikti lentelėje

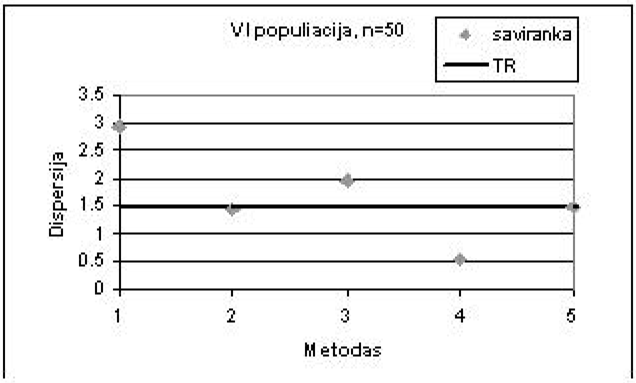
Populiacija	Imties dydis	$\widehat{cv}(\widehat{S}^2)$
I	$n = 30$	0,177736
	$n = 40$	0,143717
	$n = 50$	0,117145
	$n = 80$	0,086990
II	$n = 30$	0,138892
	$n = 40$	0,123944
	$n = 50$	0,097992
	$n = 100$	0,065856
III	$n = 30$	0,255801
	$n = 40$	0,128995
	$n = 50$	0,071350
	$n = 80$	0,072795
IV	$n = 30$	0,127082
	$n = 40$	0,112158
	$n = 50$	0,100261
	$n = 100$	0,032102

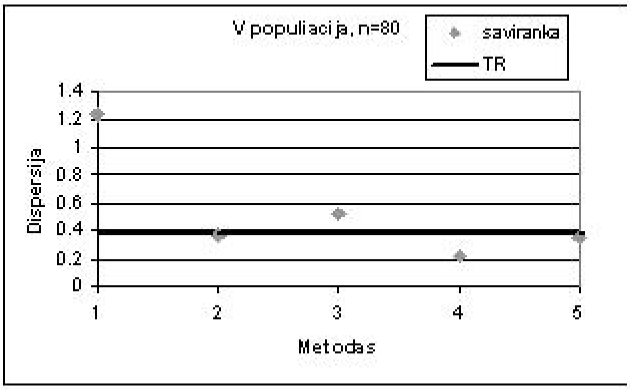
matome, kad esant didelėms populiacijoms ir mažoms imtims, dispersijos tiksliai įvertinti negalime. Tačiau praktikoje dažniausiai tenka pasitenkinti tuo, ką gavome, nes paimti didesnių imčių nėra galimybės.

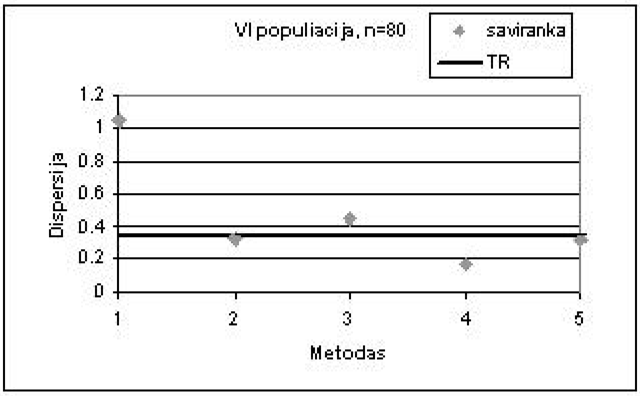

V populiacija, n=30
saviranka
TR
Dispersija
Metodas


VI populiacija, n=30
saviranka
TR
Dispersija
Metodas


V populiacija, n=50
saviranka
TR
Dispersija
Metodas


VI populiacija, n=50
saviranka
TR
Dispersija
Metodas


V populiacija, n=80
saviranka
TR
Dispersija
Metodas


VI populiacija, n=80
saviranka
TR
Dispersija
Metodas

V ir VI populiacijoje, kur nagrinėjami realūs duomenys, rezultatai iš dalies yra panašūs kaip ir generuotuose populiacijose: išlaikomas savirankos metodų išsibarstymas ir pastebima, kad geriausiai vertina antras (mastelio pakeitimo saviranka) ir penktas (grąžintinė saviranka) savirankos metodas nežiūrint į nukrypimus su atskiromis imtimis.

Pažiūrėję į variacijos koeficiento reikšmes

Populiacija	Imties dydis	$\widehat{cv}(\widehat{S}^2)$
IV	$n = 30$	0,103565
	$n = 50$	0,073793
	$n = 80$	0,041965
VI	$n = 30$	0,111487
	$n = 50$	0,090606
	$n = 80$	0,043721

matome tas pačias tendencijas kaip ir prieš tai. Tik esant $n = 80$, galime teigti, kad dispersija vertinama gana tiksliai.

Apibendrinant gautus rezultatus galime teigti, kad savirankos metodai visai neblogai vertina dispersijos įvertinio dispersiją ir skirtumai su tikrąja reikšme nėra dideli. Esant skirtingam duomenų pasiskirstymui, skirtingi savirankos metodai duoda geresnius rezultatus. Be to, pastebėjome, kad realūs duomenys labiausiai atitinka eksponentinį pasiskirstymą (to mes ir tikėjomės iš pradžių), kuriam dispersijos įvertinio dispersiją geriausiai vertina antras (mastelio pakeitimo saviranka) ir penktas (grąžintinė saviranka) savirankos metodas. Vertinant kitą populiacijos parametrą ir tą patį parametrą, bet kitoje populiacijoje, gali būti, kad kita savirankos rūšis bus efektyvesnė.

8 Išvados

Šiame darbe nagrinėjami trys baigtinės populiacijos dispersijos įverčiai, esant sluoksninėms imtims, ištirtas jų poslinkis. Vienas iš įverčių yra paslinktasis, bet iš poslinkio dydžio galime daryti išvadą, kad naudojant šį dispersijos įvertinį rezultatai iš esmės nepasikeičia. Iš praktinės dalies rezultatų galime teigti, kad savirankos metodai yra tinkami populiacijos dispersijos įverčių dispersijai vertinti. Renkant sluoksninę imtį su paprastąja negrąžintine imtimi sluoksniuose nustatyta, kad tradicinė saviranka visiems duomenims vertina prasčiausiai . Tuo tarpu negrąžintinė ir grąžintinė saviranka duoda geriausią rezultatą, kai duomenys pasiskirstę pagal normalinį dėsnį, o mastelio pakeitimo ir grąžintinė saviranka duoda geriausią rezultatą, kai duomenys pasiskirstę pagal eksponentinį dėsnį. Realiems duomenims gaunami panašūs rezultatai kaip ir eksponentinį skirstinį turinčiai populiacijai.

Literatūra

[1] R.R. Sitter, Comparing three bootstrap methods for survey data, *The Canadian Journal of Statistics*, **20**, 135-154 (1992).

[2] J.N.K. Rao, C.F.J. Wu, Resampling Inference With Complex Survey Data, *Journal of the American Statistical Association*, **83**, 231-241 (1988).

[3] R.R. Sitter,A Resampling Procedure for Complex Survey Data, *Journal of the American Statistical Association*, **87**, 755-765 (1992).

[4] J. Shao, D. Tu, *The Jackknife and Bootstrap*, Verlag, New York (1995).

[5] C. - E. Särndal, B. Swensson, J.Wretman, *Model Assisted Survey Sampling*, Springer -Verlag, New York, Berlin, Heidelberg (1992).

Priedas Nr.1

Nagrinėjami realūs duomenys

V populiacija

Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos
1	1030	39	56748	77	179000	115	408000
2	2654	40	56925	78	180000	116	432270
3	2900	41	58860	79	181000	117	447859
4	4267	42	61356	80	195050	118	460000
5	5000	43	63000	81	196000	119	465959
6	5050	44	66083	82	196000	120	541000
7	5592	45	67000	83	198927	121	553549
8	6649	46	69514	84	202396	122	572000
9	6800	47	71237	85	204240	123	572078
10	6900	48	71552	86	206000	124	573000
11	7298	49	72000	87	220590	125	582175
12	10000	50	72000	88	221000	126	588000
13	10075	51	72210	89	222000	127	590992
14	11400	52	81918	90	231619	128	659000
15	11954	53	87759	91	236500	129	708000
16	14548	54	94764	92	243000	130	774383
17	17000	55	97035	93	245000	131	885000
18	20900	56	99000	94	250000	132	958351
19	22532	57	101383	95	251657	133	990000
20	23537	58	109387	96	253000	134	1011000
21	24371	59	110000	97	260000	135	1033668
22	24529	60	122000	98	264609	136	1042000
23	27140	61	127892	99	268000	137	1054000
24	28057	62	130000	100	284948	138	1112000
25	28879	63	142365	101	292000	139	1159000
26	31574	64	154000	102	306600	140	1217731
27	34000	65	154022	103	313000	141	1277000
28	34029	66	156346	104	314744	142	1296760
29	36000	67	156699	105	318000	143	1345000
30	36667	68	157063	106	319000	144	1369000
31	39780	69	164000	107	332000	145	1395107
32	39822	70	165000	108	332000	146	1642000
33	43000	71	165000	109	334525	147	1650000
34	46236	72	165320	110	351000	148	1676000

Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos
35	48215	73	169535	111	367000	149	1768453
36	49000	74	170000	112	368000	150	1866000
37	56100	75	173375	113	380000	151	1949744
38	56552	76	176000	114	384000		

VI populiacija

Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos
1	2143	34	39316	67	173853	100	447647
2	2379	35	40383	68	179745	101	449684
3	6000	36	40500	69	184000	102	474000
4	6500	37	41380	70	187760	103	499970
5	7237	38	43000	71	195305	104	509189
6	7344	39	44000	72	195699	105	533000
7	7948	40	45718	73	200074	106	540000
8	8309	41	49447	74	204000	107	560277
9	9000	42	51000	75	210962	108	607000
10	9424	43	54885	76	217940	109	629000
11	10000	44	55905	77	233000	110	644403
12	12147	45	56066	78	236201	111	658000
13	12200	46	58397	79	246000	112	667000
14	12639	47	61775	80	281834	113	679955
15	12814	48	64984	81	286000	114	704000
16	13000	49	66000	82	293077	115	714702
17	13906	50	67964	83	298000	116	721866
18	15000	51	74000	84	306665	117	818000
19	15488	52	79956	85	321429	118	828000
20	15878	53	81605	86	328636	119	839000
21	16387	54	84273	87	330537	120	847000
22	17492	55	88138	88	334195	121	910271
23	18779	56	88577	89	335334	122	919109
24	19654	57	89199	90	346000	123	932500
25	20448	58	94406	91	347946	124	979000
26	23000	59	99000	92	353000	125	1144000
27	23196	60	107693	93	357000	126	1148000
28	23773	61	117390	94	359000	127	1150000
29	24393	62	129163	95	359299	128	1318131
30	25060	63	146068	96	366848	129	1379622

Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos	Įmonės Nr.	Pajamos
31	25822	64	150878	97	380442	130	1505000
32	26000	65	153721	98	388880	131	1593000
33	28000	66	169490	99	427422	132	1761781

Priedas Nr.2

SAS programos tekstas

```
* Kintamuju aprasymas;
LIBNAME BANDREZ 'C:\Documents and Settings\XXXX';
* Pagrindiniu kintamuju aprasymas;
* Sluoksniu skaicius;
%let sluoksniu_skaicius=1;
* Elementu kiekis populiacijoje;
%let populiacija=1;
* Populiacijos SAS duomenu aibes pavadinimas;
%let pop_lent = populiacija_158;
* Populiacijos duomenu sutartinis kodas;
* Pagal savo nusisatyta struktura atpazinimui ar tai generuoti;
* duomenys ir pagal koki desni ar realus ir pvz. kokios veiklos;
%let pop_met = 1581;
* Imties kopijos duomenu aibes sudarymui;
* Kopija vadinasi imtisXXX, kur xxx=imtis_nr;
%let imtis_nr = 1581050;
* Imties dydis per visus sluoksnius is viso;
%let imtis_viso = 50;

* Populiaciju kurimas;
* Realiu duomenu ikelimas;
data bandrez.populiacija_NNNN; input popduom; cards; duomenys
..... ; run; data BANDREZ.populiacija_norm;
* Generuojama normaline populiacija;
retain popduom; a=0; b=1;
   do i=1 to &populiacija;
   popduom=a+b*rannor(&seed_popul);
   output;
   end;
   drop a b i;
data BANDREZ.populiacija_norm;
* Duomenys paslenkami per 4, kad nebutu neigiamu;
set BANDREZ.populiacija_norm; popduom=popduom+4; run;

data BANDREZ.populiacija_exp;
* Generuojama eksponentine populiacija;
retain popduom; a=0; b=1;
```

```
    do i=1 to &populiacija;
    popduom=ranexp(&seed_popul)/b;
    output;
    end;
    drop a b i;
run;
* Surusiuojama normaline;
proc sort data=BANDREZ.populiacija_norm; by popduom; run;
* Surusiuojama eksponentine;
proc sort data=BANDREZ.populiacija_exp; by popduom; run;

%macro dalinimas_sluoksniais(pop_lent);
proc sql; select sum(popduom)/&sluoksniu_skaicius into
:sluoksnio_suma from bandrez.&pop_lent; quit;

data pop_sl; set bandrez.&pop_lent; sumwt + popduom; eil_nr + 1;
sluoksnio_nr = int(sumwt/&&sluoksnio_suma)+1;

if (sluoksnio_nr>&sluoksniu_skaicius) then
sluoksnio_nr=&sluoksniu_skaicius; drop sumwt eil_nr; run;

data bandrez.&pop_lent; set pop_sl; run; proc sql; drop table
pop_sl; quit;
%mend;

%dalinimas_sluoksniais(populiacija_NNNN);

* Metodu skaiciavimo programos vykdymas;
* Sios komandos katojamos norima kieki kartu keiciant populiacija bei imti;

%let pop_lent = populiacija_xx;
%let populiacija = 676;
%let pop_met = 158;
%let imtis_nr = 158030;
%let imtis_viso = 30;
%sl_kurimas;
%tikros_reiksmes;
%metodas1(1000);
%metodas2(1000);
%metodas3 (1000);
%metodas4 (1000);
```

```
%metodas5(1000);

* MACRO APRASYMAS ;
%macro sl_kurimas;
* Macro skirtas imties dydziams pagal kiekviena metoda;
* suskaiciuoti ir pirmines imties paeminui is populiacijos;
* sukuriant tarpines SAS duomenu aibes kiekvienam sluoksniui;
proc sql;
* Populiacijos sluoksniu ir elementu kiekviename sluoksnyje suskaiciavimas ;
create table sluoksniai as select sluoksnio_nr,count(*) as
sluoksnyje_vnt from  bandrez.&pop_lent group by sluoksnio_nr;
quit; proc sql noprint;
* Sluoksniu skaicius ir populiacijos skaicius talpinamas i kintamuosius;
select count(*) into :sluoksniu_skaicius from sluoksniai; select
sum(sluoksnyje_vnt) into :populiacija from sluoksniai; quit;

data bandrez.populiacija; set bandrez.&pop_lent; run; proc sql;
* Suskaiciuojama kiek reikia imciu is kiekvieno sluoksnio sukurdami darbine SAS duomenu aibe;
create table imciu_vnt_sluoksnyje as select sluoksnio_nr,
round(std*kiekis_sluoksnyje/std_suma*&imtis_viso) as imties_vnt
from (select sluoksnio_nr,count(*) as kiekis_sluoksnyje,
std(popduom) as std from bandrez.&pop_lent group by sluoksnio_nr),
(select sum(std_f) as std_suma from (select
sluoksnio_nr,count(*)*std(popduom) as std_f from
bandrez.populiacija group by  sluoksnio_nr)); quit;

data sluoksniai;
* Imciu is sluoksniu kiekio pridejimas i sluoksniu lentele;
merge sluoksniai imciu_vnt_sluoksnyje; by sluoksnio_nr; run;

data sluoksniai;
* Imciu kiekis 1 metodui sutampa su pirmine imtimi;
set sluoksniai; met1_m=imties_vnt; run;

data sluoksniai;
* Imciu kiekis 2 metodui;
set sluoksniai; met2_m=round(((imties_vnt-2)**2)/(imties_vnt-1));
run; data sluoksniai;
* Imciu kiekis ir kartojimo skaicius pagal metoda 3;
set sluoksniai; met3_n=(imties_vnt**2)/(sluoksnyje_vnt);
met3_k=(imties_vnt*(1-(met3_n/imties_vnt)))/
```

```
(met3_n*(1-(imties_vnt/sluoksnyje_vnt))); run; data sluoksniai;
* 3 metodo randomizacija pagal n;
set sluoksniai; if met3_n<1 then do; met3_n1=1; met3_np1=1;
met3_n2=0; met3_np2=0; end; else do; if met3_n=round(met3_n) then
do; met3_n1=met3_n; met3_np1=1; met3_n2=0; met3_np2=0; end; if
met3_n>round(met3_n) then do; met3_n1=round(met3_n);
met3_np1=round(met3_n)-met3_n+1; met3_n2=round(met3_n)+1;
met3_np2=1-met3_np1; end; if  met3_n<round(met3_n) then do;
met3_n1=round(met3_n)-1; met3_np1=round(met3_n)-met3_n;
met3_n2=round(met3_n); met3_np2=1-met3_np1; end; end; run;

data sluoksniai;
* 3 metodo randomizacija pagal k;
set sluoksniai; if met3_k=round(met3_k) then do; met3_k1=met3_k;
met3_kp1=1; met3_k2=0; met3_kp2=0; end; if  met3_k>round(met3_k)
then do; met3_k1=round(met3_k);
met3_kp1=((1/met3_k)-1/(round(met3_k)+1))/
((1/round(met3_k))-1/(round(met3_k)+1)); met3_k2=round(met3_k)+1;
met3_kp2=1-met3_kp1; end; if  met3_k<round(met3_k) then do;
met3_k2=round(met3_k); met3_kp1=((1/met3_k)-1/round(met3_k))/
((1/(round(met3_k)-1))-1/round(met3_k)); met3_k1=round(met3_k)-1;
met3_kp2=1-met3_kp1; end; run;

data sluoksniai;
* Imciu kiekis ir kartojimo skaicius pagal metoda 4;
set sluoksniai; met4_n=imties_vnt-(1-(imties_vnt/sluoksnyje_vnt));
met4_k=(sluoksnyje_vnt/imties_vnt)*(1-((sluoksnyje_vnt-imties_vnt)/
(imties_vnt*sluoksnyje_vnt))); run;

data sluoksniai;
* 4 metodo randomizacija pagal n ir k ;
set sluoksniai; if met4_n<1 then do; met4_n1=1; met4_n2=0; end;
else do; if met4_n=round(met4_n) then do; met4_n1=met4_n;
met4_n2=0; end; if  met4_n>round(met4_n) then do;
met4_n1=round(met4_n); met4_n2=round(met4_n)+1;
                          end;
if  met4_n<round(met4_n) then do; met4_n1=round(met4_n);
met4_n2=round(met4_n)-1;
                          end;
end; if met4_k=round(met4_k) then do; met4_k1=met4_k; met4_k2=0;
end; if  met4_k>round(met4_k) then do; met4_k1=round(met4_k);
```

```
met4_k2=round(met4_k)+1;
                        end;
if  met4_k<round(met4_k) then do; met4_k1=round(met4_k);
met4_k2=round(met4_k)-1;
                        end;
a_1=(met4_k1-(met4_n1/imties_vnt))/(met4_n1*(imties_vnt*met4_k1-1));
a_2=(met4_k2-(met4_n2/imties_vnt))/(met4_n2*(imties_vnt*met4_k2-1));
met4_p1=(((1-imties_vnt/sluoksnyje_vnt)/(imties_vnt*(imties_vnt-1)))-a_2)/(a_1-a_2);
met4_p2=1-met4_p1; drop a_1 a_2; run;

data sluoksniai;
* Imciu kiekis pagal metoda 5;
set sluoksniai;
met5_n=((imties_vnt-1)*sluoksnyje_vnt)/(sluoksnyje_vnt-imties_vnt);
run;

data sluoksniai;
* 5 metodo randomizacija;
set sluoksniai; if met5_n=round(met5_n) then do; met5_n1=met5_n;
met5_np1=1; met5_n2=0; met5_np2=0; end; if  met5_n>round(met5_n)
then do; met5_n1=round(met5_n);
met5_np1=((1/met5_n)-1/(round(met5_n)+1))/
((1/round(met5_n))-1/(round(met5_n)+1)); met5_n2=round(met5_n)+1;
met5_np2=1-met5_np1; end; if  met5_n<round(met5_n) then do;
met5_n2=round(met5_n); met5_np1=((1/met5_n)-1/round(met5_n))/
((1/(round(met5_n)-1))-1/round(met5_n)); met5_n1=round(met5_n)-1;
met5_np2=1-met5_np1; end; run;

proc sql;
* Darbines duomenu aibes imciu vienetams skaiciuoti istrynimas;
drop table imciu_vnt_sluoksnyje; run;

* Populiacijos padalinimas i sluoksnius;
%MACRO atskiri_sluoksniai;
%do i=1 %to &sluoksniu_skaicius;
data sluokspop&i; set bandrez.populiacija; where sluoksnio_nr=&i;
output; run;
%end;
%mend;

%MACRO imtis1;
```

```
* Imties is populiacijos paemimas kiekvienam sluoksniui;
%do i=1 %to &sluoksniu_skaicius;
proc sql noprint; select imties_vnt into :samples&i from
sluoksniai where sluoksnio_nr=&i; quit;
* Imtis negrazintine ;
proc surveyselect noprint data=sluokspop&i out=imtis_sl&i
method=srs n=&&samples&i; run;
%end;
%mend;

%MACRO imties_vidurkis;
* Apjungiame imtis vidurkio skaiciavimui;
proc sql; drop table imtis_visa; quit;
%do ii=1 %to &sluoksniu_skaicius;
proc append base=imtis_visa new=imtis_sl&ii force; run;
%end;
* Sluoksniniu imciu vidurkio suskaiciavimas;
proc sql; create table sluoks_vidurk as select
sluoksnio_nr,sum(popduom)/count(*) as imties_vidurkis from
imtis_visa group by sluoksnio_nr; quit; data sluoksniai;
* Sluoksniniu imciu vidurkio pridejimas i parametru lentele SLUOKSNIAI;
merge sluoksniai sluoks_vidurk; by sluoksnio_nr; run;
* Darbines aibes istrynimas;
proc sql; drop table sluoks_vidurk; quit;
%mend;

* Macro programu vykdymas;
%atskiri_sluoksniai;
%imtis1;
%imties_vidurkis;

* Imties kopijos padarymas velesniam nagrinejimui;
data bandrez.imtis&imtis_nr; set imtis_visa; run;
%mend;

%macro tikros_reiksmes;
* Macro skirtas imties tikrosios reiksmes bei CV paskaiciavimui;
* Populiacijos elementu skaiciaus irasymas i kintamaji;
proc sql noprint; select count(*) into :populiacija from
bandrez.&pop_lent; quit;
* Sukuriama tarpine SAS sumine duomenu aibe populiacijos lentele pagal sluoksnius;
```

```
proc sql; create table pop_sluoksniai as select sluoksnio_nr as
pop_sl,sum(popduom) as pop_suma, count(*) as N_did from
bandrez.&pop_lent group by pop_sl; quit;
* Tikrosios reiksmes formule skaidome i dalis del programos aiskumo;
* Paskaiciuojama tikrosios reiksmes formules dalis;
proc sql noprint; select sum(sumyk)/&&populiacija into :uhk_dalis1
from ( select sluoksnio_nr,(N_did/n_maz)*suma as sumyk from (
select sluoksnio_nr,sum(popduom) as suma,count(*) as n_maz from
bandrez.imtis&imtis_nr group by sluoksnio_nr),pop_sluoksniai where
pop_sl=sluoksnio_nr group by sluoksnio_nr); quit;

* Paskaiciuojama tikrosios reiksmes formules kita dalis;
proc sql noprint; select sum(sumyk)/(&&populiacija-1) into
:uhk_dalis2 from ( select sluoksnio_nr,(N_did/n_maz)*suma as sumyk
from ( select sluoksnio_nr,sum((popduom-&&uhk_dalis1)**2) as
suma,count(*) as n_maz from bandrez.imtis&imtis_nr group by
sluoksnio_nr),pop_sluoksniai where pop_sl=sluoksnio_nr group by
sluoksnio_nr); quit;

* Sukuriama tarpine SAS duomenu aibe populiacijos duomenu perskaiciavimui;
proc sql; create table uhk as select
sluoksnio_nr,popduom,((popduom-&&uhk_dalis1)**2-&&uhk_dalis2)/(&populiacija-1)
as uhk from bandrez.imtis&imtis_nr quit;

* Galutine tikroji reiksme;
proc sql; create table ds as select
sum(N_did*(N_did-n_maz)*sh/n_maz) as ds from ( select
sluoksnio_nr,sum((uhk-avgsl)**2)/(n_maz-1) as sh from uhk, (select
sluoksnio_nr as sl_nr1,avg(uhk) as avgsl,count(*) as n_maz from
uhk group by sl_nr1) where sluoksnio_nr=sl_nr1 group by
sluoksnio_nr),pop_sluoksniai where pop_sl=sluoksnio_nr; quit;

* cv s kvadrato paskaiciavimas;
proc sql noprint; select sum(nn*suma)/(&populiacija-1) into
:s_skvt from ( select a.sluoksnio_nr as sl_nrc,
sluoksnyje_vnt/imties_vnt as nn, (popduom-&&uhk_dalis1)**2 as suma
from bandrez.imtis&imtis_nr a,sluoksniai b where
b.sluoksnio_nr=sl_nrc) ; quit;

* Tarpiniu SAS duomenu aibiu istrynimas;
proc sql; drop table pop_sluoksniai; drop table uhk; quit;
```

```
* SAS duomenu aibeje kaupiamos tikroji reiksme ir CV, identifikuojant pagal imties nr;
proc sql noprint; insert into bandrez.tikros_reiksmes select
&imtis_nr, ds , sqrt(ds)/&&s_skvt from ds ; quit;
%mend;

* METODAS 1 ;

%MACRO metodas1(ivertiniu_vnt);
%do i=1 %to &ivertiniu_vnt;
%if &i>1 %then %do;
proc sql; drop table temp_visi_sl; quit;
%end;
%do ii=1 %to &sluoksniu_skaicius;
proc sql noprint; select met1_m into :samples&ii from sluoksniai
where sluoksnio_nr=&ii; quit;

data temp; pt=int(ranuni(0)*nobs)+1; set imtis_sl&ii point=pt
nobs=nobs; k+1; IF k > &&samples&ii THEN STOP; drop k; run; proc
append base=temp_visi_sl new=temp force; run;
%end;

proc sql; create table ivertinys1 as select &i as ivert_nr,
sum((sluoksnyje_vnt/met1_m)*suma)/(&populiacija-1) as ivertinys
from ( select a.sluoksnio_nr,sum((popduom-miu)**2) as suma,
met1_m, sluoksnyje_vnt from temp_visi_sl a,sluoksniai b, (select
sum((sluoksnyje_vnt/met1_m)*suma)/&populiacija as miu from (
select a.sluoksnio_nr as sluoksnio_nr ,sluoksnyje_vnt , met1_m,
sum(a.popduom)as suma from temp_visi_sl a,sluoksniai b where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met1_m )) where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met1_m ); quit;

proc append base=ivertiniai1_visi new=ivertinys1 force; run;
%end;
proc sql; insert into bandrez.variacijos select &pop_met, '1' as
metodas,&sluoksniu_skaicius, &imtis_nr, &imtis_viso,
&ivertiniu_vnt as ivertiniu_vnt,
sum((ivertinys-vidurkis)**2)/kiekis as variacija from
ivertiniai1_visi, (select avg(ivertinys) as vidurkis, count(*) as
```

```
kiekis from ivertiniai1_visi) ; quit;

proc datasets; delete ivertiniai1_visi; delete ivertinys1; delete
temp_visi_sl; delete temp; run;
%mend;

* METODAS 2 ;

%MACRO metodas2(ivertiniu_vnt);
%do i=1 %to &ivertiniu_vnt;
%if &i>1 %then %do;
proc sql; drop table temp_visi_sl; quit;
%end;
%do ii=1 %to &sluoksniu_skaicius;
proc sql noprint; select met2_m into :samples&ii from sluoksniai
where sluoksnio_nr=&ii; quit; data temp;
pt=int(ranuni(&seed_resampl+&i+&ii)*nobs)+1; set imtis_sl&ii
point=pt nobs=nobs; k+1; IF k > &&samples&ii THEN STOP; drop k;
run; proc append base=temp_visi_sl new=temp force; run;
%end;
proc sql noprint; create table y_bangeles as select
a.sluoksnio_nr,
imties_vidurkis+(sqrt(met2_m)/(sqrt(imties_vnt-1))*
sqrt(1-(imties_vnt/sluoksnyje_vnt)))* (popduom-imties_vidurkis) as
y_bangele from temp_visi_sl a,sluoksniai b where
a.sluoksnio_nr=b.sluoksnio_nr; quit;

proc sql; create table ivertinys2 as select &i as
ivert_nr,sum((sluoksnyje_vnt/met2_m)*suma)/(&populiacija-1) as
ivertinys from ( select a.sluoksnio_nr,met2_m,sluoksnyje_vnt,
sum((y_bangele-miu)**2) as suma from y_bangeles a,sluoksniai b,
(select sum((sluoksnyje_vnt/met2_m)*suma)/&populiacija as miu from
(select a.sluoksnio_nr as sluoksnio_nr ,b.sluoksnyje_vnt,met2_m,
sum(a.y_bangele)as suma from y_bangeles a,sluoksniai b where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met2_m )) where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,met2_m,sluoksnyje_vnt ); quit;
 proc append base=ivertiniai2_visi new=ivertinys2 force;
run;
%end;
```

```
proc sql; insert into bandrez.variacijos select  &pop_met, '2' as
metodas,&sluoksniu_skaicius, &imtis_nr,&imtis_viso, &ivertiniu_vnt
as ivertiniu_vnt, sum((ivertinys-vidurkis)**2)/(kiekis-1) as
variacija from ivertiniai2_visi, (select avg(ivertinys) as
vidurkis, count(*) as kiekis from ivertiniai2_visi) ; quit; proc
datasets; delete ivertiniai2_visi; delete ivertinys2; delete temp;
delete temp_visi_sl; delete y_bangeles; run;
%mend;

* METODAS 3 ;

%MACRO metodas3 (ivertiniu_vnt);
data po_kiek_kartu; set sluoksniai;
pirmas_n=round(&ivertiniu_vnt*met3_np1);
antras_n=&ivertiniu_vnt-pirmas_n;
pirmas_k=round(&ivertiniu_vnt*met3_kp1);
antras_k=&ivertiniu_vnt-pirmas_k; run;
%do ll=1 %to &sluoksniu_skaicius;
proc sql noprint;
* Po kiek imti ?;
select met3_n1 into :n1&&ll  from po_kiek_kartu where
sluoksnio_nr=&ll; select met3_n2 into :n2&&ll  from po_kiek_kartu
where sluoksnio_nr=&ll; select met3_k1 into :k1&&ll  from
po_kiek_kartu where sluoksnio_nr=&ll; select met3_k2 into :k2&&ll
from po_kiek_kartu where sluoksnio_nr=&ll; select pirmas_n into
:nk&ll from po_kiek_kartu where sluoksnio_nr=&ll; select pirmas_k
into :kk&ll from po_kiek_kartu where sluoksnio_nr=&ll; quit;
%end;

%do i=1 %to &ivertiniu_vnt;
%if &i>1 %then %do;
proc sql; drop table temp_visi_sl; quit;
%end;
%do ii=1 %to &sluoksniu_skaicius;
%if &i<=%sysevalf(&&nk&ii) %then
%let po_kiek=%sysevalf(&&n1&ii);
%else
%let po_kiek=%sysevalf(&&n2&ii);
%if &i<=%sysevalf(&&kk&ii) %then
%let kartoti=%sysevalf(&&k1&ii);
%else
```

```
%let kartoti=%sysevalf(&&k2&ii);

%do ik=1 %to &kartoti;
proc surveyselect noprint data=imtis_sl&ii out=temp method=srs
n=&po_kiek; run; proc append base=temp_visi_sl new=temp force;
run;
%end;
%end;
proc sql; create table ivertinys3 as select &i as
ivert_nr,sum((sluoksnyje_vnt/kiekis)*suma)/(&populiacija-1) as
ivertinys from ( select a.sluoksnio_nr,sum((popduom-miu)**2) as
suma, sluoksnyje_vnt,count(*) as kiekis from temp_visi_sl
a,sluoksniai b, (select
sum((sluoksnyje_vnt/kiekis)*suma)/&populiacija as miu from (select
a.sluoksnio_nr as sluoksnio_nr ,sluoksnyje_vnt, sum(a.popduom)as
suma,count(*) as kiekis from temp_visi_sl a,sluoksniai b where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,sluoksnyje_vnt )) where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,sluoksnyje_vnt ); quit;

 proc append base=ivertiniai3_visi new=ivertinys3 force;
run;
%end;
proc sql; insert into bandrez.variacijos select &pop_met, '3' as
metodas,&sluoksniu_skaicius, &imtis_nr,&imtis_viso, &ivertiniu_vnt
as ivertiniu_vnt, sum((ivertinys-vidurkis)**2)/kiekis as variacija
from ivertiniai3_visi, (select avg(ivertinys) as vidurkis,
count(*) as kiekis from ivertiniai3_visi) ; quit; proc datasets;
delete ivertiniai3_visi; delete ivertinys3; delete temp_visi_sl;
delete temp; delete po_kiek_kartu; run;
%mend;

* METODAS 4;
%MACRO metodas4(ivertiniu_vnt);

%do ll=1 %to &sluoksniu_skaicius;
proc sql noprint; select round(met4_n1) into :n41&ll from
sluoksniai where sluoksnio_nr=&ll; select round(met4_n2) into
:n42&ll  from sluoksniai where sluoksnio_nr=&ll;
```

```
select round(met4_k1) into :k41&ll from sluoksniai where
sluoksnio_nr=&ll; select round(met4_k2) into :k42&ll  from
sluoksniai where sluoksnio_nr=&ll;

* Tikimybe;
select round(met4_p1) into :p41&ll from sluoksniai where
sluoksnio_nr=&ll; select round(met4_p2) into :p42&ll  from
sluoksniai where sluoksnio_nr=&ll;

quit;
* Pseudo1 popupliacijos sukurimas;
%do ik=1 %to %sysevalf(&&k41&ll);
proc append base=sl_pseudo1&ll new=imtis_sl&ll force; run;
%end;
* Pseudo2 popupliacijos sukurimas;
%do ik=1 %to %sysevalf(&&k42&ll);
proc append base=sl_pseudo2&ll new=imtis_sl&ll force; run;
%end;
%end;

%do i=1 %to &ivertiniu_vnt;
%if &i>1 %then %do;
proc sql; drop table temp_visi_sl; quit;
%end;

%do ll=1 %to &sluoksniu_skaicius;
%let pop_keitimas=%sysevalf(&ivertiniu_vnt)*%sysevalf(&&p41&ll);
%if &i<=&pop_keitimas %then %do;
* Jei i<= n1 tai is pseudo1;
proc surveyselect noprint data=sl_pseudo1&ll out=temp method=srs
n= %sysevalf(&&n41&ll);
run;
%end;
%else %do;
* Jei i>n1 tai is pseudo2;
proc surveyselect noprint data=sl_pseudo2&ll out=temp method=srs
n= %sysevalf(&&n42&ll);
run;
%end;
proc append base=temp_visi_sl new=temp force; run;
%end;
```

```
proc sql; create table ivertinys4 as select &i as
ivert_nr,sum((sluoksnyje_vnt/met4_n)*suma)/(&populiacija-1) as
ivertinys from ( select a.sluoksnio_nr,sum((popduom-miu)**2) as
suma, met4_n, b.sluoksnyje_vnt from temp_visi_sl a,sluoksniai b,
(select sum((sluoksnyje_vnt/met4_n)*suma)/&populiacija as miu from
(select a.sluoksnio_nr as sluoksnio_nr ,b.sluoksnyje_vnt,
sum(a.popduom)as suma,met4_n from temp_visi_sl a,sluoksniai b
where a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met4_n )) where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met4_n ); quit;
 proc append base=ivertiniai4_visi new=ivertinys4 force;
run;
%end;

proc sql; insert into bandrez.variacijos select &pop_met, '4' as
metodas,&sluoksniu_skaicius, &imtis_nr,&imtis_viso, &ivertiniu_vnt
as ivertiniu_vnt, sum((ivertinys-vidurkis)**2)/kiekis as variacija
from ivertiniai4_visi, (select avg(ivertinys) as vidurkis,
count(*) as kiekis from ivertiniai4_visi) ; quit; proc datasets;
%do ll=1 %to &sluoksniu_skaicius;
delete sl_pseudo1&ll; delete sl_pseudo2&ll;
%end;
delete ivertiniai4_visi; delete ivertinys4; delete temp; delete
temp_visi_sl; run;
%mend;

* METODAS 5 ;

%MACRO metodas5(ivertiniu_vnt);
%do ll=1 %to &sluoksniu_skaicius;
proc sql noprint; select round(met5_n1) into :n51&&ll  from
sluoksniai where sluoksnio_nr=&ll; select round(met5_n2) into
:n52&&ll  from sluoksniai where sluoksnio_nr=&ll; select
round(met5_np1) into :p51&&ll  from sluoksniai where
sluoksnio_nr=&ll; select round(met5_np2) into :p52&&ll  from
sluoksniai where sluoksnio_nr=&ll;
%end;

%do i=1 %to &ivertiniu_vnt;
```

```
%if &i>1 %then %do;
proc sql; drop table temp_visi_sl; quit;
%end;

%do ii=1 %to &sluoksniu_skaicius;
%let pop_keitimas=%sysevalf(&ivertiniu_vnt)*%sysevalf(&&p51&ii);
%if &i<=&pop_keitimas %then %do;
data temp; pt=int(ranuni(0)*nobs)+1; set imtis_sl&ii point=pt
nobs=nobs; k+1;
IF k > %sysevalf(&&n51&ii) THEN STOP;
drop k; run;
%end;
%if &i>&pop_keitimas %then %do;
data temp; pt=int(ranuni(0)*nobs)+1; set imtis_sl&ii point=pt
nobs=nobs; k+1;
IF k > %sysevalf(&&n52&ii) THEN STOP;
drop k; run;
%end;
proc append base=temp_visi_sl new=temp force; run;
%end;

proc sql; create table ivertinys5 as select &i as
ivert_nr,sum((sluoksnyje_vnt/met5_n)*suma)/(&populiacija-1) as
ivertinys from ( select a.sluoksnio_nr,sum((popduom-miu)**2) as
suma, met5_n, b.sluoksnyje_vnt from temp_visi_sl a,sluoksniai b,
(select sum((sluoksnyje_vnt/met5_n)*suma)/&populiacija as miu from
( select a.sluoksnio_nr as sluoksnio_nr ,b.sluoksnyje_vnt,
sum(a.popduom)as suma,met5_n from temp_visi_sl a,sluoksniai b
where a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,b.sluoksnyje_vnt,met5_n )) where
a.sluoksnio_nr=b.sluoksnio_nr group by
a.sluoksnio_nr,sluoksnyje_vnt,met5_n ); quit;
 proc append base=ivertiniai5_visi new=ivertinys5 force;
run;
%end;
proc sql; insert into bandrez.variacijos select &pop_met, '5' as
metodas,&sluoksniu_skaicius, &imtis_nr,&imtis_viso, &ivertiniu_vnt
as ivertiniu_vnt, sum((ivertinys-vidurkis)**2)/(kiekis-1) as
variacija from ivertiniai5_visi, (select avg(ivertinys) as
vidurkis, count(*) as kiekis from ivertiniai5_visi) ; quit;
```

```
proc datasets; delete ivertiniai5_visi; delete ivertinys5; delete
temp_visi_sl; delete temp; run;
%mend;
```