VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

Dalius PUMPUTIS

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI, GAUTI NAUDOJANT PAPILDOMĄ INFORMACIJĄ

Daktaro disertacija

Fiziniai mokslai, matematika (01P)

Vilnius VGTU LEIDYKLA TECHNIKA 2008

Disertacija rengta 2004–2008 metais Matematikos ir informatikos institute.

Darbo vadovas

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

Konsultantas

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

http://leidykla.vgtu.lt
VGTU leidyklos TECHNIKA 1576-M mokslo literatūros knyga

ISBN 978-9955-28-388-1

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
INSTITUTE OF MATHEMATICS AND INFORMATICS

Dalius PUMPUTIS

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

Doctoral Dissertation

Physical Sciences, Mathematics (01P)

VGTU LEIDYKLA TECHNIKA

Vilnius 2008

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

Scientific Supervisor

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

Consultant

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

Reziumė

Disertacijoje nagrinėjamos papildomos informacijos panaudojimo galimybės konstruojant baigtinės populiacijos sumos, dispersijos ir kovariacijos įvertinius, bei sluoksniuojant baigtines populiacijas.

Pirmiausia darbe sprendžiamas populiacijų sluoksniavimo uždavinys, kai tyrimo kintamojo skirstinys yra asimetrinis. Pasiūlomas naujas – pataisytasis geometrinis – sluoksniavimo metodas. Šis metodas modeliuojant lyginamas su trimis kitais žinomais metodais: kvadratinės šaknies iš skirstinio dažnio, geometriniu ir laipsninio sluoksniavimo metodu. Modeliavimo rezultatai rodo, kad vidutiniškai asimetrinėms populiacijoms geriausiai tinka laipsninio sluoksniavimo metodas, o ypač asimetrinėms populiacijoms geriausias yra pataisytasis geometrinis sluoksniavimas.

Toliau nagrinėjami baigtinės populiacijos sumos kalibruotieji įvertiniai, sukonstruoti taikant skirtingas atstumo funkcijas. Modeliuojant tiriama šių įvertinių kokybė. Sukonstruoti nauji populiacijos kovariacijos kalibruotieji įvertiniai, naudojantys vieną, dvi ir tris svorių sistemas. Šie įvertiniai konstruojami pasirenkant skirtingas kalibravimo lygtis. Remiantis modeliais pagrįstų įvertinių teorija, čia taip pat sukonstruojamas pataisytasis tiesiniu regresiniu modeliu pagrįstas kalibruotasis populiacijos kovariacijos įvertinys. Modeliuojant įvertiniai lyginami tarpusavyje ir su standartiniais atitinkamų parametrų įvertiniais. Kalibruotieji įvertiniai yra kur kas tikslesni, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė. Disertacijoje taip pat sprendžiama sukonstruotų įvertinių dispersijų vertinimo problema. Tai – techniškai sudėtingas uždavinys, nes daugeliu atvejų kalibruotųjų svorių negalima užrašyti išreikštiniu pavidalu.

Abstract

The dissertation analyzes how to incorporate auxiliary information into the estimation of the finite population total, variance, covariance, and how to use it for the stratification of finite populations.

First of all, the problem of efficient stratification in the case of skewed population is considered. A new adjusted geometric stratification method is introduced. This method is compared by simulation with the cumulative root frequency method, the geometric method, and the power method. The simulation results show that in most cases considered the power method is the most efficient one, but the adjusted geometric stratification method outperforms all the methods in the case of highly skewed populations.

The calibrated estimators of finite population total, constructed using different distance functions, are considered. The quality of such estimators is analyzed by simulation. The new calibrated estimators of the finite population covariance (variance) are derived, using one or more weighting systems. Applying the model calibration theory, we construct here an adjusted linear regression model-assisted and calibrated estimator of the population covariance. The estimators derived are compared by simulation with the standard estimators of the respective parameters. The calibrated estimators of the population covariance are more efficient compared to the straight estimators provided the auxiliary variables are well correlated with the study variables. The problem of estimation of the variance of the constructed estimators is considered in the dissertation as well. This is a technically complicated problem, because an explicit solution of calibration equations does not exist in many cases.

Žymėjimai

$\mathcal{U}$	baigtinė populiacija, sudaryta iš N elementų;
$\mathbf{s}$	tikimybinė imtis;
N	populiacijos dydis;
n	imties dydis;
H	sluoksnių skaičius;
U_h	populiacijos h-asis sluoksnis;
N_h	h-ojo sluoksnio dydis;
n_h	imties dydis h-ajame sluoksnyje;
$Y_1, Y_2, \ldots, Y_N$	nepriklausomi atsitiktiniai dydžiai;
y, z, $y^{(1)}, y^{(2)}, \ldots, y^{(M)}$	tyrimo kintamieji;
$x^{(1)}, x^{(2)}, \ldots, x^{(J)}$	papildomi kintamieji;
x_y, a	tyrimo kintamojo y papildomi kintamieji;
x_z, b	tyrimo kintamojo z papildomi kintamieji;
$\mathbf{x}_k$	k-ojo populiacijos elemento papildomų kintamųjų reikšmių vektorius stulpelis;
$\mathbf{x}'_k$	transponuotasis vektorius $\mathbf{x}_k$;
$\mathbf{y}_k$	k-ojo populiacijos elemento tyrimo kintamųjų reikšmių vektorius stulpelis;
$p(\mathbf{s})$	imties $\mathbf{s}$ išrinkimo tikimybė;
π_k	k-ojo populiacijos elemento priklausymo imčiai tikimybė;
π_{kl}	dviejų populiacijos elementų k ir l priklausymo imčiai tikimybė;
d_k	k-ojo populiacijos elemento imties plano svoris;
w_k	k-ojo imties elemento kalibruotasis svoris;
w_{kl}	elementų poros (k, l) kalibruotasis svoris;
q_k	k-ojo populiacijos elemento laisvai pasirenkamas svoris;
q_{kl}	elementų poros (k, l) laisvai pasirenkamas svoris;
$\bar{y}$	kintamojo y imties vidurkis;
s_y^2	kintamojo y imties dispersija;
t_y	kintamojo y suma;
μ_y	kintamojo y vidurkis;
R	dviejų populiacijos sumų t_y ir t_z santykis;

S_y^2	kintamojo y dispersija;
$Cov(y, z)$	kintamųjų y ir z kovariacija;
$\rho(y, z)$	kintamųjų y ir z koreliacijos koeficientas;
$F_y(t)$	kintamojo y pasiskirstymo funkcija;
Q_α	kintamojo y α lygmens kvantilis;
ac	populiacijos asimetrijos koeficientas;
$\widehat{\theta}$	parametro θ įvertinys;
$POSL(\widehat{\theta})$	įvertinio $\widehat{\theta}$ poslinkis;
$AVar(\widehat{\theta})$	įvertinio $\widehat{\theta}$ apytikslė dispersija;
$Var(\widehat{\theta})$	įvertinio $\widehat{\theta}$ dispersija;
$VKP(\widehat{\theta})$	įvertinio $\widehat{\theta}$ vidutinė kvadratinė paklaida;
$cv(\widehat{\theta})$	įvertinio $\widehat{\theta}$ variacijos koeficientas;
$L, L^\star$	atstumo funkcijos (jos gali būti numeruojamos, tada prirašomas atitinkamas indeksas);
$Cum\sqrt{f}$	šaknies iš f taisyklė;
NPK	įvertiniai, naudojantys papildomus kintamuosius.

Turinys

Padėka

Nuoširdžiai dėkoju darbo vadovams doc. dr. Aleksandrui Plikusui ir prof. habil. dr. Romanui Januškevičiui už vadovavimą disertacijai, konsultacijas ir joms skirtą laiką; doc. dr. Danutei Krapavickaitei ir doc. dr. (HP) Marijui Radavičiui už vertingas pastabas ir patarimus disertacijos struktūros ir terminų klausimais; prof. Gunarui Kuldorfui (Gunnar Kulldorf, Umeå universitetas, Švedija), prof. Danieliui Torburnui (Daniel Thorburn, Stokholmo universitetas, Švedija), doc. dr. Danutei Krapavickaitei ir doc. dr. Aleksandrui Plikusui už stažuotes Umeå ir Stokholmo universitetuose, kurių metu buvo gauti pagrindiniai disertacijos rezultatai, sudarytas disertacijos literatūros sąrašas ir užmegzti dalykiniai ryšiai su švedų imčių metodų teorijos specialistais; kolegoms iš Vilniaus pedagoginio universiteto Algebros ir statistikos katedros už moralinį palaikymą; Zitai Manstavičienei už sutvarkytą rankraščio kalbą.

Įvadas

Tiriamoji problema

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

Darbo aktualumas

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

Daugelyje mokslo darbų pabrėžiama žinomos papildomos informacijos svarba konstruojant populiacijos parametrų įvertinius – jei papildomos informacijos kintamieji gerai koreliuoja su tyrimo kintamaisiais, tai pasitelkus tinkamus įvertinius, kurie naudoja papildomus kintamuosius, galima gauti tikslesnius įverčius. Ypač svarbus yra Deville ir Särndal [21] straipsnis, kuriame pristatoma nauja baigtinės populiacijos sumos įvertinių, naudojančių papildomus kintamuosius, klasė – *kalibruotieji įvertiniai*. Šie įvertiniai vis plačiau taikomi valstybinėje statistikoje.

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

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

Tyrimų objektas

Darbo tyrimų objektas yra:

- baigtinėse populiacijose apibrėžtų parametrų vertinimas naudojant papildomą informaciją;
- modeliais pagrįsti baigtinės populiacijos kovariacijos įvertiniai;
- baigtinių populiacijų sluoksniavimas.

Darbo tikslas ir uždaviniai

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

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

2. Sukonstruoti baigtinės populiacijos dispersijos ir kovariacijos kalibruotus įvertinius, naudojančius keletą svorių sistemų.
3. Sukonstruoti gautųjų kalibruotų įvertinių dispersijos įvertinius.
4. Remiantis matematiniu modeliavimu kalibruotus įvertinius palyginti su standartiniais atitinkamų parametrų įvertiniais.
5. Modeliuojant palyginti autoriaus [A1] sukonstruotus kalibruotus kovariacijos įvertinius su Wu ir Sitter [100] pasiūlytu tiesiniu regresiniu modeliu pagrįstu kalibruotu baigtinės populiacijos kovariacijos įvertiniu.
6. Modifikuoti tiesiniu regresiniu modeliu pagrįstą kalibruotą baigtinės populiacijos kovariacijos įvertinį kiekvienam tyrimo kintamajam naudojant atskirus papildomus kintamuosius.
7. Modifikuoti Gunning ir Horgan [35] pasiūlytą geometrinį sluoksniavimo metodą, darant prielaidą, kad populiacijos skirstinys yra eksponentinis.

Tyrimų metodai

Darbe taikomi šie metodai: analizinis, tikimybinis ir eksperimentinis. Įrodant disertacijos teiginius, taikyti Lagranžo daugiklių, Teiloro skleidimo eilute, atsitiktinių dydžių skaitinių charakteristikų skaičiavimo metodai ir matricų diferencijavimo taisyklės. Išvedant pataisytąjį tiesiniu regresiniu modeliu pagrįstą baigtinės populiacijos kovariacijos įvertinį, remtasi modeliais pagrįstų įvertinių teorija. Sprendžiant optimaliųjų sluoksnių ribų radimo uždavinį, nagrinėti šaknies iš f, geometrinis ir laipsnių sluoksniavimo metodai. Išvestas pataisytasis geometrinis populiacijų sluoksniavimo metodas yra pagrįstas populiacijos skirstiniu. Atliekant matematinį modeliavimą, naudotasi matematinių uždavinių sprendimo paketu Matlab [60].

Mokslinis darbo naujumas

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

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

Pataisytasis tiesiniu regresiniu modeliu pagrįstas kovariacijos įvertinys – taip pat naujas rezultatas. Jis skiriasi nuo Wu ir Sitter [100] pasiūlyto įvertinio tuo, kad jį konstruojant kiekvienam tyrimo kintamajam buvo naudotas atskiras papildomas kintamasis.

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

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

Autoriaus dalyvavimas mokslinėse programose

Autorius dalyvavo:

- Vykdant Lietuvos valstybinio mokslo ir studijų fondo finansuotą projektą „Baigtinės populiacijos parametrų vertinimas turint papildomos informacijos“ (2007 03 12 – 2007 12 30).
- 2007 m. gegužės 31 d. ir birželio 1 d. Helsinkyje vykusiuose daugiapakopio modeliavimo kursuose. Kursų, rengtų pagal Baltijos ir Skandinavijos šalių imčių metodų teorijos mokslinio bendradarbiavimo programą, organizatoriai: Helsinkio universitetas, Suomijos statistikos departamentas, Suomijos statistikų draugija.

Autorius stažavosi šiose užsienio mokslo įstaigose:

- Umeå universitete (Švedija) nuo 2006 m. vasario 28 d. iki kovo 31 d. Stažuotės metu vietiniame Matematinės statistikos katedros seminare perskaityta paskaita „Calibrated estimators under different distance measures“. Stažuotė finansuota iš projekto „Baltic-Nordic network co-operation on education and research in survey statistics“ lėšų.
- Umeå universitete (Švedija) nuo 2007 m. vasario 1 d. iki vasario 28 d. Stažuotės metu Matematinės statistikos katedros seminare perskaityta paskaita „Calibrated estimators of the finite population covariance“. Stažuotė finansuota iš projekto „Baltic-Nordic network co-operation on expanding the capacity in survey sampling and related statistical applications“ lėšų.

- Stokholmo universitete (Švedija) nuo 2008 m. sausio 8 d. iki vasario 6 d. Stažuotės metu Matematinės statistikos katedros seminare perskaityta paskaita „Estimation of the finite population covariance in the presence of auxiliary information“. Stažuotė finansuota iš projekto „Baltic-Nordic network co-operation on education and research in survey statistics“ lėšų.

Ginamieji disertacijos teiginiai

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

Darbo rezultatų aprobavimas

Pagrindiniai disertacijos rezultatai paskelbti šiuose moksliniuose žurnaluose: *Acta Applicandae Mathematicae*, *Statistics in Transition*, *Lietuvos matematikos rinkinyje* (spec. numeryje).

Tarpiniai disertacijos rezultatai pristatyti šiose mokslinėse konferencijose ir seminaruose:

- D. Pumputis, Kalibruoti įvertiniai esant skirtingoms atstumo funkcijoms, *Lietuvos matematikų draugijos XLVI konferencija*, VU, Vilnius, birželio 15–16 d., 2005.
- A. Plikusas, D. Pumputis, Laiko eilučių šuolių išlyginimas, *Lietuvos matematikų draugijos XLVI konferencija*, VU, Vilnius, birželio 15–16 d., 2005.

- D. Pumputis, Calibrated estimators under different distance measures, *Workshop on Survey Sampling Theory and Methodology*, Vilnius, June 17–21, 2005.
- A. Plikusas, D. Pumputis, Composite estimators for the sample and frame changing points, *Workshop on Survey Sampling Theory and Methodology*, Vilnius, June 17–21, 2005.
- D. Pumputis, Calibrated estimators under different distance measures, *Seminar in Survey Sampling*, University of Umeå, Umeå, Sweden, March 22, 2006.
- A. Plikusas, D. Pumputis, Kalibruoti baigtinės populiacijos dispersijos įvertiniai, esant skirtingoms atstumo funkcijoms, *Lietuvos matematikų draugijos XLVII konferencija*, KTU, Kaunas, birželio 20–21 d., 2006.
- A. Plikusas, D. Pumputis, Calibrated estimators of the population covariance, *9th International Vilnius Conference on Probability Theory and Mathematical Statistics*, Vilnius, June 25–30, 2006.
- A. Plikusas, D. Pumputis, Calibrated estimators of finite population covariance, *Workshop on Survey Sampling Theory and Methodology*, Ventspils, Latvia, August 24–28, 2006.
- D. Pumputis, Calibrated estimators of the finite population covariance, *Seminar in Survey Sampling*, University of Umeå, Umeå, Sweden, February 27, 2007.
- D. Pumputis, On the estimation of the variance of calibrated estimators of the population covariance, *Second Baltic-Nordic Conference on Survey Sampling*, Kuusamo, Finland, June 2–7, 2007.
- D. Pumputis, Populiacijos, turinčios eksponentinį skirstinį, sluoksniavimas, *Lietuvos matematikų draugijos XLVIII konferencija*, VGTU, Vilnius, birželio 27–28 d., 2007.
- A. Plikusas, D. Krapavickaitė, D. Pumputis, Estimation of variance for the calibrated estimators of the finite population covariance, *56th Session of the International Statistical Institute (ISI)*, Lisboa Congress Centre, Lisboa, Portugal, August 22–29, 2007.
- A. Plikusas, D. Pumputis, Estimation of the finite population covariance, *8th International Conference Computer Data Analysis and Modeling: Complex Stochastic Data and Systems*, Minsk, Belarus, September 11–15, 2007.

- D. Pumputis, Estimation of the finite population covariance in the presence of auxiliary information, *Seminar in Survey Sampling*, University of Stockholm, Stockholm, Sweden, January 30, 2008.
- D. Pumputis, Calibrated and model-calibrated estimators of the finite population covariance, *22nd Nordic Conference on Mathematical Statistics*, Vilnius, June 16–19, 2008.
- D. Pumputis, Kalibruoti ir modeliais pagrįsti baigtinės populiacijos kovariacijos įvertiniai, *Lietuvos matematikų draugijos XLIX konferencija*, Kaunas, birželio 25–26 d., 2008.
- D. Pumputis, Adjusted model-calibrated estimators of the finite population covariance, *Workshop on Survey Sampling Theory and Methodology*, Kuressaare, Estonia, August 25–29, 2008.
- D. Pumputis, Baigtinės populiacijos parametrų statistiniai įvertiniai, gauti naudojant papildomą informaciją, *Imčių teorijos seminaras*, Matematikos ir informatikos institutas, Vilnius, rugsėjo 24 d., 2008.

Disertacijos struktūra

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

Pirmajame skyriuje apžvelgiami kitų autorių parašyti darbai nagrinėjama tema.

Be literatūros apžvalgos sluoksniavimo tematika, antrajame skyriuje taip pat pristatomi pirmieji autoriaus rezultatai, gauti sprendžiant populiacijų sluoksniavimo uždavinį, kai tyrimo kintamojo skirstinys yra asimetrinis. Pasiūlomas naujas – pataisytasis geometrinis – sluoksniavimo metodas. Šis metodas modeliuojant lyginamas su trimis kitais žinomais kvadratinės šaknies iš skirstinio dažnio, geometriniu ir laipsninio sluoksniavimo metodais. Modeliavimo rezultatai rodo, kad vidutiniškai asimetrinėms populiacijoms geriausiai tinka laipsninio sluoksniavimo metodas, o ypač asimetrinėms populiacijoms geriausias yra pataisytasis geometrinis sluoksniavimas.

Trečiojo skyriaus pirmajame poskyryje konstruojami baigtinės populiacijos sumos kalibruoti įvertiniai, naudojant skirtingas atstumo funkcijas ir papildomą informaciją. Čia pateikiamos sukonstruotų įvertinių apytikslės dispersijos išraiškos. Atliekamas matematinis modeliavimas, kurio metu tarpusavyje lyginami sukonstruoti įvertiniai. Antrajame skyriaus poskyryje randamos kalibruotų svorių išraiškos, vertinant baigtinės populiacijos dispersiją ir kovariaciją. Tyrimui pasiūlomos skirtingos kalibravimo lygtys. Pristatomi kalibruoti baigtinės populiacijos kovariacijos įvertiniai, naudojantys kelias svorių sistemas. Remiantis matematiniu modeliavimu, sukonstruoti įvertiniai lyginami tarpusavyje ir su standartiniais įvertiniais. Trečiajame skyriaus poskyryje sprendžiamas sukonstruotų kovariacijos įvertinių dispersijų vertinimo uždavinys. Paskutiniame poskyryje nagrinėjamas kalibruotų baigtinės populiacijos kovariacijos įvertinių, naudojančių dvi svorių sistemas, ir Wu bei Sitter [100] pasiūlyto tiesiniu regresiniu modeliu pagrįsto kalibruoto kovariacijos įvertinio palyginimo uždavinys. Čia taip pat pateikiamas pataisytasis tiesiniu regresiniu modeliu pagrįstas kovariacijos įvertinys.

1

Papildomą informaciją naudojantys baigtinės populiacijos parametrų įvertiniai

1.1. Papildoma informacija

Baigtinių populiacijų statistikoje yra nagrinėjamos fiksuotos populiacijos $\mathcal{U}= \{u_1, u_2, \dots, u_N\}$, turinčios baigtinį skaičių elementų. Tiriami baigtinėje populiacijoje apibrėžti ir realias reikšmes įgyjantys (tyrimo) kintamieji. Tokių kintamųjų, pavyzdžiui, kintamojo y, reikšmes žymėsime $y_1, y_2, \dots, y_N$ ir laikysime pastoviomis. Tipiškas statistikos uždavinys – tai populiacijos skaitinių charakteristikų vertinimas. Populiacijos skaitinių charakteristikų pavyzdžiai: kintamojo reikšmių suma, dalis populiacijos elementų, turinčių tam tikrą savybę, populiacijos vidurkis, populiacijos standartinis nuokrypis ir t.t. Skaitinė populiacijos charakteristika – tai populiacijos kintamojo reikšmių funkcija, kitaip dar vadinama *populiacijos parametru*. Turint baigtinę populiaciją, kurią sudaro N elementų, ir joje apibrėžtą kintamąjį $y : y_1, y_2, \dots, y_N$, galima sudaryti įvairiausių jos parametrų $\theta = \theta(y_1, y_2, \dots, y_N)$, apibūdinančių šią populiaciją. Viena iš galimybių tirti populiaciją ir vertinti mus dominančius parametrus – rinkti duomenis tik iš dalies populiacijos elementų. Tą populiacijos dalį vadinsime *imtimi* ir žymėsime $\mathrm{s} = \{u_{i_1}, u_{i_2}, \dots, u_{i_n}\}$. Čia n – imties elementų skaičius, vadinamas *imties dydžiu*. Imties elementų kintamojo y reikšmes žymėsime $y_{i_1}, y_{i_2}, \dots, y_{i_n}$.

Tokie tyrimai, kai renkami ir nagrinėjami imties elementų duomenys, siekiant padaryti išvadas apie visą populiaciją, vadinami *imčių tyrimais*. Disertacijoje nagrinėsime tikimybines imtis, t. y. tokias, kurios gaunamos taikant tam tikrą imties išrinkimo procedūrą, tenkinančią šiuos reikalavimus [56]:

1) nusakomos visos galimos skirtingos imtys $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_v$;

2) kiekvienai galimai imčiai $\mathbf{s}_i$ priskiriama žinoma imties išrinkimo tikimybė $p(\mathbf{s}_i) > 0$, $i = 1, 2, \dots, v$, taip, kad $\sum_{k=1}^{v} p(\mathbf{s}_k) = 1$;

3) kiekvienas populiacijos elementas priklauso bent vienai galimai imčiai;

4) imtis išrenkama taikant tokią atsitiktinę procedūrą, kai kiekviena iš galimų imčių $\mathbf{s}_i$ išrenkama su nurodyta tikimybe $p(\mathbf{s}_i)$, $i = 1, 2, \dots, v$.

Visos galimos imtys $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_v$ ir tikimybės, su kuriomis jos gali būti išrinktos $p(\mathbf{s}_1), p(\mathbf{s}_2), \dots, p(\mathbf{s}_v)$, apibrėžia tikimybinį skirstinį, kuris yra vadinamas *imties planu*. Dydis $\pi_k = P(\mathbf{s} : k \in \mathbf{s})$ vadinamas k-ojo populiacijos elemento, o $\pi_{kl} = P(\mathbf{s} : k\&l \in \mathbf{s})$ – dviejų populiacijos elementų k ir l priklausymo imčiai tikimybėmis. Populiacijos k-ojo elemento priklausymo imčiai tikimybę π_k ir imties išrinkimo tikimybes sieja lygybė:

$$\pi_k = \sum_{i:k \in \mathbf{s}_i} p(\mathbf{s}_i).$$

Dydžius, atvirkštinius elementų priklausymo imčiai tikimybėms, vadinsime *imties plano svoriais* ir žymėsime $d_k = \frac{1}{\pi_k}$.

Pati paprasčiausia tikimybinė imtis yra *paprastoji atsitiktinė negrąžintinė imtis*. Tai tokia n skirtingų elementų imtis iš baigtinės N dydžio populiacijos, kai bet kuris n skirtingų elementų rinkinys turi vienodą tikimybę būti išrinktas. Šiuo atveju kiekvienos n dydžio imties išrinkimo tikimybė yra $p(\mathbf{s}) = \frac{1}{C_N^n}$, o $\pi_k = \frac{n}{N}$, $k = 1, 2, \dots, N$.

Išrinkus imtį ir surinkus imties elementų duomenis, populiacijos parametrų vertinimui pasitelkiamos tam tikros imties elementų funkcijos, kurios vadinamos parametrų įvertiniais. *Įvertinys* – tai statistika, kuri naudojama populiacijos parametrui įvertinti, o *įvertis* – tai įvertinio reikšmė, gauta iš konkrečios imties. Parametro θ įvertinį žymėsime $\widehat{\theta} = \widehat{\theta}(y_{i_1}, y_{i_2}, \dots, y_{i_n})$.

Turint kelis skirtingus populiacijos parametro θ įvertinius, „geresnis“ bus tas, kuris bus arčiau populiacijos parametro θ, t. y. kuo bus mažesnė įverčių, gautų iš skirtingų imčių, sklaida apie tikrąją populiacijos parametro θ reikšmę.

Skirtumas $POSL(\widehat{\theta}) = E\widehat{\theta} - \theta$ vadinamas įvertinio $\widehat{\theta}$ *poslinkiu*. Kai poslinkis $POSL(\widehat{\theta}) = 0$, tai įvertinys $\widehat{\theta}$ vadinamas *nepaslinktuoju*. Baigtinės populiacijos atveju įvertinio $\widehat{\theta}$ nepaslinktumas reiškia, kad paėmus visas skirtingas n dydžio imtis ir apskaičiavus statistikos $\widehat{\theta}$ realizacijas, visų šių realizacijų vidurkis imties plano tikimybių atžvilgiu yra lygus θ (šis vidurkis sutampa su aritmetiniu vidurkiu tada, kai imtys renkamos su vienodomis tikimybėmis $p(\mathbf{s}) = \frac{1}{C_N^n}$). Įvertis gali būti nepaslinktasis, nors nė viena jo realizacija nesutampa su vertinamuoju parametru. Tačiau ne visada galima sukonstruoti nepaslinktąjį įvertinį. Poslinkis – vienas iš įvertinio $\widehat{\theta}$ tikslumo matų. Kiti populiariausi tikslumo matai yra įvertinio $\widehat{\theta}$:

- *dispersija*
$$Var(\widehat{\theta}) = E(\widehat{\theta} - E\widehat{\theta})^2,$$
apibūdinanti įvertinio $\widehat{\theta}$ sklaidą apie jo vidutinę reikšmę;

- *vidutinė kvadratinė paklaida*
$$VKP(\widehat{\theta}) = E(\widehat{\theta} - \theta)^2,$$
parodanti įvertinio $\widehat{\theta}$ sklaidos lygį apie tikrąją parametro θ reikšmę;

- *standartinė paklaida*
$$SP(\widehat{\theta}) = \sqrt{Var(\widehat{\theta})},$$
atspindinti tą patį, ką ir įvertinio dispersija;

- *santykinė standartinė paklaida (variacijos koeficientas)*
$$cv(\widehat{\theta}) = \frac{\sqrt{Var(\widehat{\theta})}}{E\widehat{\theta}}.$$
Kuo variacijos koeficientas mažesnis, tuo įvertinys tikslesnis.

Parametro θ įvertinys $\widehat{\theta}$ vadinamas *suderintuoju*, jei jo vidutinė kvadratinė paklaida artėja į nulį ($VKP(\widehat{\theta}) \to 0$), kai populiacija ir imtis tam tikru būdu neaprėžtai didėja. Populiacijos ir imties didėjimo būdai aptariami literatūroje, kurioje nagrinėjamos asimptotinės įvertinių savybės.

Pagal savo prigimtį populiacijos parametrų įvertiniai skirstomi į:

- pagrįstus tik imties planu. Jie naudoja imties elementų tyrimo kintamojo reikšmes ir imties plano tikimybes;

- naudojančius imties elementų tyrimo kintamojo reikšmes, imties plano tikimybes ir papildomą informaciją.

Papildoma informacija yra plati sąvoka, aprėpianti:

- populiacijoje apibrėžtus kintamuosius $x^{(1)}, x^{(2)}, \dots, x^{(J)}$, kurių reikšmės yra žinomos kiekvienam išrinktos imties elementui arba visiems populiacijos elementams. Šiuos kintamuosius vadinsime *papildomais kintamaisiais*;

- kurių nors populiacijos kintamųjų skaitines charakteristikas (vidurkį, dispersiją ir t.t.), kurios yra žinomos, bet pačios tų kintamųjų reikšmės nėra žinomos visiems populiacijos elementams;

- kitą žinomą informaciją apie populiaciją.

Papildoma informacija gali būti žinoma, pavyzdžiui, iš ankstesnio gyventojų surašymo, iš statistinių įmonių registrų (čia yra įvairios ekonominės informacijos) ar kitų šaltinių. Šioje disertacijoje, sprendžiant įvade suformuluotus uždavinius, bus naudojami papildomi kintamieji, kurių reikšmių vektorių k-ajam populiacijos elementui žymėsime $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$. Trumpiniu NPK žymėsime įvertinius, kurie naudoja papildomus kintamuosius.

Papildomi kintamieji galėtų būti skirstomi į *kokybinius* ir *kiekybinius*. Tarkime, kad atliekame imčių tyrimą, kurio metu siekiame įvertinti visos Lietuvos dirbančiųjų atlyginimų vidurkį. Taigi tiriame visos Lietuvos dirbančiųjų populiaciją ir joje apibrėžtą tyrimo kintamąjį y – dirbančiojo atlyginimą. Konstruojant minėto vidurkio NPK įvertinius, kokybiniu papildomu kintamuoju x galėtų būti dirbančiojo lytis ar amžiaus grupė, tuo tarpu kintamasis, rodantis atlyginimų dydį praeitais metais, yra kiekybinis.

Imčių tyrimuose papildoma informacija gali būti naudojama:

- imties plano sudarymui;
- parametrų įvertinių konstravimui;
- ir imties plano sudarymui, ir įvertinių konstravimui.

Kitame poskyryje nagrinėsime baigtinės populiacijos sumos ir vidurkio įvertinius, naudojančius papildomą informaciją.

1.2. Įvertiniai, naudojantys papildomą informaciją

1.2.1. Populiacijos sumos *NPK* įvertiniai

Nagrinėkime baigtinę populiaciją $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$, sudarytą iš N elementų. Tegul y, įgyjantis reikšmes $y_1, y_2, \ldots, y_N$, yra tyrimo kintamasis.

Nagrinėsime vieną iš paprasčiausių parametrų – populiacijos sumą

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

Horvitz ir Thompson [46] 1952 metais pasiūlė gana universalų nepaslinktąjį populiacijos sumos įvertinį

$$\hat{t}_{y\pi} = \sum_{k \in \mathbf{s}} \frac{y_k}{\pi_k}, \tag{1.1}$$

tinkantį vertinti bet kokio imties plano atveju. Šis įvertinys, autorių garbei pavadintas *Horvico ir Tompsono įvertiniu*, tapo pagrindu daugelio vėlesnių mokslo darbų. Vieni iš jų – nagrinėjantys sudėtingesnius, papildomus kintamuosius naudojančius įvertinius.

Tarkime, kad turime papildomą kintamąjį x, įgyjantį reikšmes $x_1, x_2, \ldots, x_N$. Tegul yra žinoma suma $t_x = \sum_{k=1}^{N} x_k$ ir imties elementų papildomo kintamojo x reikšmės. Įvertinys

$$\hat{t}_{yr} = \frac{\hat{t}_{y\pi}}{\hat{t}_{x\pi}} t_x$$

vadinamas *santykiniu populiacijos sumos įvertiniu*, čia $\hat{t}_{y\pi}$ ir $\hat{t}_{x\pi}$ yra atitinkamai kintamųjų y ir x Horvico ir Tompsono įvertiniai. Santykinis sumos įvertinys yra tikslesnis, jei koreliacija tarp y ir x yra didelė.

Toliau tarkime, kad turime ne vieną, o J papildomų kintamųjų

$$x^{(1)}, x^{(2)}, \ldots, x^{(J)}.$$

Pažymėkime $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ populiacijos k-ojo elemento papildomų kintamųjų reikšmių vektorių. Tegul tyrimo kintamąjį y ir papildomus kintamuosius $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$ sieja tiesinis regresinis modelis ξ. Čia y yra priklausomas, o $x^{(1)}, \ldots, x^{(J)}$ – nepriklausomi kintamieji. Modelio ξ savybės:

1. Tyrimo kintamojo reikšmės $y_1, y_2, \ldots, y_N$ yra nepriklausomų atsitiktinių dydžių $Y_1, Y_2, \ldots, Y_N$ realizacijos.
2. $E_\xi(Y_k) = \sum_{j=1}^{J} \beta_j x_{jk}$.
3. $V_\xi(Y_k) = \sigma_k^2$.

Čia E_ξ ir V_ξ yra atitinkamai atsitiktinio dydžio Y_k vidurkis ir dispersija modelio ξ atžvilgiu, o $\beta_1, \ldots, \beta_J$ ir $\sigma_1^2, \ldots, \sigma_N^2$ – modelio parametrai.

Naudojantis tiesiniu regresiniu modeliu, apibrėžiamas *apibendrintasis regresinis populiacijos sumos įvertinys*

$$\hat{t}_{yr} = \hat{t}_{y\pi} + \sum_{j=1}^{J} \widehat{\beta}_j (t_{x^{(j)}} - \hat{t}_{x^{(j)}\pi}), \tag{1.2}$$

čia $\hat{t}_{y\pi} = \sum_{k \in \mathbf{s}} y_k/\pi_k$, $\hat{t}_{x^{(j)}\pi} = \sum_{k \in \mathbf{s}} x_{jk}/\pi_k$ yra atitinkamai tyrimo kintamojo sumos t_y ir papildomo kintamojo $x^{(j)}$ sumos $t_{x^{(j)}}$ Horvico ir Tompsono įvertiniai.

Regresijos koeficientų įverčiai $\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_J$ yra J-mačio vektoriaus

$$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_J)' = \Big(\sum_{k\in\mathbf{s}} \mathbf{x}_k\mathbf{x}_k'/\sigma_k^2\pi_k\Big)^{-1} \sum_{k\in\mathbf{s}} \mathbf{x}_k y_k/\sigma_k^2\pi_k$$

komponentės. Dviejų papildomų kintamųjų $x^{(1)}$ ir $x^{(2)}$ atveju, kai $x_{1k} = 1$, $k = 1, 2, \ldots, N$, (1.2) įvertinys vadinamas tiesiog *regresiniu populiacijos sumos įvertiniu*.

Iš apibendrinto regresinio įvertinio analizinės išraiškos matome, kad pastarasis įvertinys yra gautas prie Horvico ir Tompsono sumos įvertinio pridėjus pataisą $\sum_{j=1}^{J} \widehat{\beta}_j(t_{x^{(j)}} - \hat{t}_{x^{(j)}\pi})$, kurios koreliacija su įvertinio $\hat{t}_{y\pi}$ paklaidomis dažnai yra neigiama. Apibendrinto regresinio įvertinio efektyvumas, palyginti su Horvico ir Tompsono įvertiniu, labai priklauso nuo to, kaip tiesinis regresinis modelis atskleidžia tikruosius kintamųjų ryšius. Juk nebūtinai kintamųjų priklausomybė turi būti tiesinė [23]. Apibrėžtą apibendrintąjį regresinį įvertinį ir jo savybes nagrinėjo Isaki ir Fuller [48], Robinson ir Särndal [79], Cassel, Särndal ir Wretman [8] ir kt. Viename iš naujesnių straipsnių Fuller [33] ne tik apžvelgia pagrindines apibendrintojo regresinio įvertinio savybes, bet ir nagrinėja (1.2) įvertinio dispersijos vertinimo alternatyvas. Lehtonen, Särndal ir Veijanen [59] lygina sintetinį, sudėtinį [85] ir apibendrintąjį regresinį įvertinius, vertindami sumas mažose populiacijos srityse. Daug statistinių tyrimų reikia įvertinti parametrus ne tik visoje populiacijoje, bet ir atskirose jos dalyse, kurios yra vadinamos *sritimis*. Vertinant parametrus srityse, iš kiekvienos srities renkama imtis. Jei srities imtis turi vos kelis elementus, tai ta sritis vadinama *maža sritimi*. Įverčių tikslumas, vertinant sumas mažose srityse, labai priklauso nuo naudojamo įvertinio ir nuo modelio tinkamumo duomenims aprašyti. Jei modelis gerai aprašo tikruosius kintamųjų ryšius, tai skirtumas tarp sintetinio ir apibendrintojo regresinio įvertinio yra nežymus. Priešingu atveju atsiranda didelis skirtumas – sintetinis įvertinys turi kur kas didesnę dispersiją.

Straipsnio [43] 4 ir 5 skyreliuose užrašomas ir nagrinėjamas populiacijos sumos apibendrintasis regresinis įvertinys ir jo dispersija dviejų fazių imties atveju. *Dviejų fazių ėmimas* – tai toks tikimybinis ėmimas, kai iš išrinktos imties renkama kita, mažesnė imtis.

1.2.2. Populiacijos vidurkio *NPK* įvertiniai

Publikacijose dažnai pasitaiko dar vienas populiarus parametras – populiacijos vidurkis μ_y. Išraišką (1.2) padauginus iš $\frac{1}{N}$ gaunamas tyrimo kintamojo y vidurkio apibendrintasis regresinis įvertinys $\widehat{\mu}_{yr} = \frac{1}{N}\hat{t}_{yr}$. Straipsnyje [19] nagrinėjamas paprastosios atsitiktinės imties atvejis. Teoriškai ir empiriškai tarpusavyje lyginami keli vidurkio regresinio įvertinio, gaunamo iš apibendrintojo regresinio vidurkio įvertinio tiesiniame regresiniame modelyje ξ naudojant tik vieną papildomą kintamąjį, dispersijos įvertiniai.

Kaip alternatyvą regresiniam vidurkio įvertiniui paprastosios atsitiktinės imties ir vieno papildomo kintamojo x atveju pateikė Srivastava [102]. Populiacijos vidurkiui vertinti jis pasiūlė ištisą įvertinių klasę

$$\tilde{y}_h = \bar{y}h(\bar{x}/\mu_x), \tag{1.3}$$

čia μ_x – kintamojo x vidurkis, $\bar{x} = \frac{1}{n}\sum_{k\in\mathbf{s}} x_k$, $\bar{y} = \frac{1}{n}\sum_{k\in\mathbf{s}} y_k$, o $h(\cdot)$ – tokia parametrinė funkcija, kuri tenkina tam tikras reguliarumo sąlygas ir $h(1) = 1$. Buvo įrodyta, kad pateiktų įvertinių asimptotinės vidutinės kvadratinės paklaidos apatinis rėžis sutampa su regresinio įvertinio asimptotinės vidutinės kvadratinės paklaidos apatiniu rėžiu. Vėlesniame darbe [103] įvertinių klasė (1.3) buvo praplėsta. Gauti įvertiniai, kurių pavidalas yra

$$\tilde{y}_t = \bar{y}t(u, v), \tag{1.4}$$

čia $t(u, v)$ – aprėžta ir tolydinė dviejų realiųjų kintamųjų u ir v funkcija, kuri tenkina lygybę $t(1, 1) = 1$, turi tolydžias ir aprėžtas pirmosios bei antrosios eilės išvestines. Išsprendus (1.4) įvertinių vidutinės kvadratinės paklaidos minimizavimo uždavinį, paaiškėja, kad praplėstoje klasėje yra tokių vidurkio įvertinių, kurie yra tikslesni už kiekvieną (1.3) klasės įvertinį.

1.3. Neatsakymai į apklausas

Beveik kiekviename tyrime kai kurie į imtį įtraukti populiacijos elementai nepateikia savo duomenų. Tokio reiškinio priežasčių yra daug: sunkiai pasiekiamas imties elementas; apklausos metu imties elementas nebuvo namuose; dėl asmeninių ar kitokių priežasčių į imtį išrinktas elementas atsisakė ar negalėjo atsakyti į kai kuriuos klausimus; apklausos anketa dėl blogo transportavimo ar kitų veiksnių negrįžo į tyrimų centrą ir t.t.

Kaip elgtis esant neatsakymams į apklausas – vienas iš aktualiausių šiuolaikinės imčių teorijos uždavinių. Taikant standartinius įvertinius ir esant neatsakymams į apklausą, galima gauti paslinktuosius įverčius. Todėl tobulinami teoriniai parametrų vertinimo esant neatsakymams metodai, galintys sumažinti neatsakymų įtaką įverčiams. Daromos įvairios prielaidos apie neatsakymų prigimtį, naudojamasi papildoma informacija, taikomi trūkstamų kintamųjų reikšmių įrašymo metodai.

Kaip sumažinti neigiamą neatsakymų į apklausas poveikį tyrimų rezultatams, aptariama straipsniuose [5] ir [33]. Čia populiacijos sumos ar vidurkio vertinimui naudojami regresinio tipo įvertiniai.

Įvertinių tyrinėjimui neatsakymų į apklausas atveju Bethlehem [5] įveda populiacijos elementų atsakymo į apklausas tikimybes $\kappa_k, k = 1 \dots N$. Straipsnyje parodoma, kad apibendrintojo regresinio populiacijos vidurkio įvertinio poslinkis gali būti sumažintas, jei duomenis aprašantis modelis parinktas taip, kad kuo labiau paaiškintų tyrimo kintamojo elgesį.

Regresiniai populiacijos sumos ir vidurkio įvertiniai tiriami Fuller [33] straipsnyje, nagrinėjant atvirkštinius atsakymų tikimybėms dydžius

$$\kappa_i^{-1} = g(\mathbf{z}_i; \boldsymbol{\zeta}^0),$$

čia $\mathbf{z}_i$ – kintamųjų vektorius, kuris gali būti stebimas tiek atsakiusiems, tiek neatsakiusiems populiacijos elementams; $g(\mathbf{z}_i; \boldsymbol{\zeta})$ yra tolydi funkcija (atsakymo tikimybių funkcija), turinti tolydžias pirmosios ir antrosios eilės išvestines atviroje aibėje, kurioje su visais $\mathbf{z}_i$ yra tikroji parametro $\boldsymbol{\zeta}$ reikšmė $\boldsymbol{\zeta}^0$.

Tegul δ_i yra indikatorinis kintamasis, kurio reikšmė lygi 1, jei i-asis populiacijos elementas atsakė į apklausa, $\delta_i = 0$, jei neatsakė. Tada naudojant vektorių $(\delta_i, \mathbf{z}_i)$ yra įvertinamas atsakymo tikimybių funkcijos parametras $\boldsymbol{\zeta}$, tuo pačiu ir dydžiai κ_i^{-1}. Taigi $\widehat{\kappa}_i^{-1} = g(\mathbf{z}_i; \widehat{\boldsymbol{\zeta}})$, o regresijos koeficientams vertinti siūlomas šis įvertinys

$$\tilde{\mathbf{B}} = \Big(\sum_{k \in \mathbf{s}} \mathbf{x}_k \mathbf{x}_k' p_k^{-1} \widehat{\kappa}_k^{-1} \delta_k \Big)^{-1} \sum_{k \in \mathbf{s}} \mathbf{x}_k y_k p_k^{-1} \widehat{\kappa}_k^{-1} \delta_k,$$

čia p_k – k-ojo populiacijos elemento išrinkimo į imtį tikimybė. Pavyzdžiui, paprastojo atsitiktinio grąžintinio ėmimo atveju populiacijos elementų išrinkimo į imtį tikimybės yra $p_k = \frac{1}{N}$.

Yra daug kitų būdų ir metodų, sumažinančių neigiamą neatsakymų į apklausas poveikį tyrimo rezultatams. Vėliau grįšime prie šios problemos sprendimo nagrinėdami kalibruotus populiacijos sumos įvertinius.

1.4. Sudėtingesnių populiacijos parametrų vertinimas naudojant papildomą informaciją

Sudėtingesnių populiacijos parametrų, kaip kad populiacijoje apibrėžto tyrimo kintamojo dispersija, kovariacija, kvantilis ir pan., vertinimas turint papildomos informacijos nėra plačiai aptariamas publikacijose.

Singh ir Joarder [97] nagrinėja baigtinėje populiacijoje $\mathcal{U}$ apibrėžto tyrimo kintamojo y dispersijos $S_y^2 = (N-1)^{-1}\sum_{i=1}^N (y_i - \mu_y)^2$ vertinimą neatsakymų į apklausas atveju ir turint papildomą kintamąjį x. Imties planas – paprastasis atsitiktinis negrąžintinis. Buvo nagrinėtos šios situacijos:

1. Tariama, kad populiacijos elementų, kurie neatsakė į apklausą, tyrimo ir papildomo kintamųjų reikšmės yra nežinomos, bet papildomo kintamojo dispersija $S_x^2 = (N-1)^{-1}\sum_{i=1}^N (x_i - \mu_x)^2$ yra žinoma. Šiuo atveju siūlomas toks tyrimo kintamojo y dispersijos įvertinys:

$$\widehat{S}_{1y}^2 = s_y^{*2}\frac{S_x^2}{s_x^{*2}},$$

čia

$$s_y^{*2} = (n-b-1)^{-1}\sum_{i=1}^{n-b}(y_i - \bar{y}^*)^2, \quad s_x^{*2} = (n-b-1)^{-1}\sum_{i=1}^{n-b}(x_i - \bar{x}^*)^2,$$

$$\bar{y}^* = (n-b)^{-1}\sum_{i=1}^{n-b} y_i, \quad \bar{x}^* = (n-b)^{-1}\sum_{i=1}^{n-b} x_i,$$

b – neatsakiusių į apklausą elementų skaičius.

2. Daroma prielaida, kad yra nežinomos tik tyrimo kintamojo y reikšmės. Taigi žinomos visos papildomo kintamojo x reikšmės ir jo dispersija S_x^2. Šiuo atveju tyrimo kintamojo y dispersijos įvertinys apibrėžiamas taip:

$$\widehat{S}_{2y}^2 = s_y^{*2}\frac{S_x^2}{s_x^2},$$

čia

$$s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \bar{x} = n^{-1}\sum_{i=1}^{n} x_i.$$

3. Nagrinėjamas atvejis, kai neatsakiusių į apklausą populiacijos elementų tyrimo kintamojo reikšmės yra nežinomos, o papildomo kintamojo reikšmės – žinomos visiems imties elementams, bet dispersija S_x^2 nežinoma. Dispersijos vertinimui pasiūlytas šis įvertinys:

$$\widehat{S}_{3y}^2 = s_y^{*2}\frac{s_x^2}{s_x^{*2}}.$$

Straipsnyje [97] išvestos apytikslės minėtų įvertinių poslinkio, vidutinės kvadratinės paklaidos ir jos įvertinio išraiškos. Taip pat pateikti ir kiti tyrimo kintamojo

y dispersijos įvertiniai.

Publikacijoje [76] nagrinėjamas baigtinėje populiacijoje apibrėžto kintamojo y pasiskirstymo funkcijos

$$F(t) = N^{-1} \sum_{k \in U} \Delta(t - y_k)$$

vertinimas turint papildomus kintamuosius. Čia $\Delta(a) = 1$, jei $a \geq 0$, ir $\Delta(a) = 0$ – priešingu atveju.

Vienas iš klasikinių pasiskirstymo funkcijos įvertinių yra

$$\widehat{F}(t) = \sum_{k \in \mathbf{s}} \pi_k^{-1} \Delta(t - y_k) \Big/ \sum_{k \in \mathbf{s}} \pi_k^{-1}.$$

Šis įvertinys nenaudoja papildomų kintamųjų. Tarkime, kad turime vieną papildomą kintamąjį x, įgyjantį reikšmes $x_1, x_2, \ldots, x_N$.

Straipsnyje [76] pateikiami du pasiskirstymo funkcijos įvertiniai: santykinis ir skirtuminis. Jie yra tiesiogiai gaunami iš žinomų rezultatų apie sumas ar vidurkius, laikant $\Delta(t - y_k)$ ir $\Delta(t - \widehat{R}x_k)$ kintamaisiais y ir x, čia

$$\widehat{R} = \Big(\sum_{k \in \mathbf{s}} y_k / \pi_k \Big) \Big(\sum_{k \in \mathbf{s}} x_k / \pi_k \Big)^{-1}$$

yra santykio $R = \sum_{k \in U} y_k \big/ \sum_{k \in U} x_k$ įvertinys. Taigi *santykinis* $F(t)$ įvertinys yra

$$\widehat{F}_r(t) = N^{-1} \frac{\sum_{k \in \mathbf{s}} \pi_k^{-1} \Delta(t - y_k)}{\sum_{k \in \mathbf{s}} \pi_k^{-1} \Delta(t - \widehat{R}x_k)} \Big\{ \sum_{k \in U} \Delta(t - \widehat{R}x_k) \Big\}. \tag{1.5}$$

Galima pastebėti, kad santykinis įvertinys sutampa su tikrąja pasiskirstymo funkcija, jei kintamasis y yra proporcingas kintamajam x. Šiuo atveju įvertinio dispersija yra lygi 0. Tai rodo, kad sukonstruotas įvertinys gali duoti apčiuopiamų rezultatų efektyvumo prasme, jei kintamasis y yra apytiksliai proporcingas kintamajam x.

Skirtuminis pasiskirstymo funkcijos įvertinys, turintis tą pačią savybę, kaip ir (1.5), apibrėžiamas taip

$$\begin{aligned}\widehat{F}_d(t) =& N^{-1}\Bigg[\sum_{k\in\mathbf{s}}\pi_k^{-1}\Delta(t-y_k)\\ &+\Bigg\{\sum_{k\in U}\Delta(t-\widehat{R}x_k)-\sum_{k\in\mathbf{s}}\pi_k^{-1}\Delta(t-\widehat{R}x_k)\Bigg\}\Bigg]\end{aligned} \tag{1.6}$$

Abu pasiskirstymo funkcijos įvertiniai yra asimptotiškai nepaslinkti.

Turint įvertintą pasiskirstymo funkciją ir papildomus kintamuosius, lengvai galima pereiti prie tyrimo kintamojo kvantilių vertinimo uždavinio. Baigtinėje populiacijoje apibrėžto kintamojo y α lygmens *kvantiliu* vadiname

$$Q_\alpha = inf\{t : F(t) \geq \alpha\}. \tag{1.7}$$

Taigi, norint įvertinti kvantilius, reikia mokėti įvertinti kintamųjų pasiskirstymo funkcijas. Pasinaudoję (1.5) ir (1.6) formulėmis ir padarę prielaidą, kad $\widehat{F}_r(t)$, $\widehat{F}_d(t)$ – monotoniškos nemažėjančios funkcijos, gauname du tyrimo kintamojo kvantilių įvertinius:

$$\widehat{Q}_{r\alpha} = inf\{t : \widehat{F}_r(t) \geq \alpha\}$$

ir

$$\widehat{Q}_{d\alpha} = inf\{t : \widehat{F}_d(t) \geq \alpha\}$$

Jei tyrimo kintamojo y ir papildomo kintamojo x koreliacija yra didelė, tai pasiskirstymo funkcijos įvertiniai (1.5), (1.6) ir jais naudojantis gauti kvantilių įvertiniai turi mažesnes dispersijas nei tradiciniai tik imties planu pagrįsti minėtų parametrų įvertiniai.

Rueda ir Arcos [80] nagrinėja pataisytąjį santykinį α lygmens kvantilio įvertinį

$$\widehat{Q}_{adj.\alpha} = \widehat{Q}_{y\alpha}\left(\frac{Q_{x\alpha}}{\widehat{Q}_{x\alpha}}\right)^{\beta}, \tag{1.8}$$

čia $\widehat{Q}_{y\alpha},\ \widehat{Q}_{x\alpha}$ – atitinkamai tyrimo ir papildomo kintamųjų standartiniai imties planu pagrįsti α lygmens kvantilių įvertiniai, skaičius $\beta \in R$.

Pateiktas įvertinys yra paslinktas, bet kai kurioms parametro β reikšmėms turintis mažesnę vidutinę kvadratinę paklaidą nei tradicinis santykinis įvertinys. Straipsnio autoriai gauna optimalios β reikšmės išraišką. Tai reiškia, kad su ta reikšme įvertinio (1.8) vidutinė kvadratinė paklaida yra minimizuojama.

Atskiru atveju $\alpha = 0{,}5$ lygmens kvantilis vadinamas *mediana*. Šio parametro vertinimo nuosekli analizė aptariama [57] straipsnyje.

Baigtinės populiacijos parametrų, kurių pavidalas yra

$$H^{\star} = \sum_{k\in U} h(y_k),$$

vertinimas turint papildomos informacijos nagrinėjamas [75] publikacijoje. Čia $h(\cdot)$ – tai bet kuri funkcija, apibrėžta realiųjų skaičių aibėje. Atskiru atveju, kai $h(y) = N^{-1}\Delta(t-y)$, parametras $H^{\star}$ yra kintamojo y pasiskirstymo funkcija $H^{\star} = F(t) = N^{-1}\sum_{k\in U}\Delta(t-y_k)$.

1.5. Sudėtingų įvertinių dispersijų vertinimas

Turint kelis tam tikro populiacijos parametro įvertinius, iškyla tų įvertinių palyginimo uždavinys. Norint išspręsti šį uždavinį, reikia mokėti apskaičiuoti įvertinių dispersijas. Ne visiems įvertiniams galima rasti tikslią dispersijos išraišką. Tada skaičiuojama apytikslė dispersija arba ji vertinama. Šiame poskyryje aptarsime du populiarius sudėtingų įvertinių dispersijų vertinimo metodus.

1.5.1. Teiloro ištiesinimo[1] metodas apytikslei dispersijai rasti

Tarkime, kad baigtinėje populiacijoje $\mathcal{U}$ turime M apibrėžtų tyrimo kintamųjų $y^{(1)},\ldots,y^{(M)}$, čia $y^{(j)}$ įgyjamos reikšmės yra $y_{j1},\ldots,y_{jN}$. Nagrinėkime tokį baigtinės populiacijos $\mathcal{U}$ parametrą, kuris gali būti išreiškiamas kaip funkcija, priklausanti nuo tyrimo kintamųjų $y^{(1)},\ldots,y^{(M)}$ sumų:

$$\theta = f(t_1,\ldots,t_M), \quad t_j = \sum_{k=1}^{N} y_{jk}.$$

Vienas iš šio parametro standartinių įvertinių yra

$$\widehat{\theta}_{\pi} = f(\hat{t}_{1\pi},\ldots,\hat{t}_{M\pi}),$$

čia $\hat{t}_{j\pi} = \sum_{k\in\mathbf{s}}\frac{y_{jk}}{\pi_k}$ – tyrimo kintamojo $y^{(j)}$ sumos t_j Horvico ir Tompsono įvertinys.

Įvertinys $\widehat{\theta}_{\pi}$ yra funkcija, priklausanti nuo kintamųjų $\hat{t}_{1\pi},\ldots,\hat{t}_{M\pi}$. Rasime jo apytikslę dispersiją Teiloro ištiesinimo metodu kelių kintamųjų funkcijoms. Ši technika leidžia aproksimuoti netiesinį įvertinį $\widehat{\theta}_{\pi}$ pseudoįvertiniu $\widehat{\theta}_0$, kuris yra gaunamas skleidžiant funkciją f Teiloro eilute taške $(\hat{t}_{1\pi},\ldots,\hat{t}_{M\pi}) = (t_1,\ldots,t_M)$

[1] Angl. Taylor linearization.

ir toliau imant tik pirmosios eilės narius:

$$\widehat{\theta}_0 = \theta + \sum_{j=1}^{M} a_j(\hat{t}_{j\pi} - t_j),$$

čia $a_j = \frac{\partial f}{\partial \hat{t}_{j\pi}}\Big|_{(\hat{t}_{1\pi},\ldots,\hat{t}_{M\pi})=(t_1,\ldots,t_M)}$.

Toliau apytikslę dispersiją skaičiuojame remdamiesi gerai žinoma populiacijos sumos Horvico ir Tompsono įvertinio dispersijos išraiška [14]:

$$\begin{aligned} AVar(\widehat{\theta}) &= Var\left(\sum_{j=1}^{M} a_j \hat{t}_{j\pi}\right) = Var\left(\sum_{k\in \mathrm{s}} \sum_{j=1}^{M} a_j \frac{y_{jk}}{\pi_k}\right) \\ &= \sum_{k\in U}\sum_{l\in U} (\pi_{kl} - \pi_k\pi_l)\breve{u}_k\breve{u}_l, \end{aligned} \tag{1.9}$$

čia $\breve{u}_k = \sum_{j=1}^{M} a_j \frac{y_{jk}}{\pi_k}$.

Šios apytikslės dispersijos įvertinys yra

$$\widehat{AVar}(\widehat{\theta}) = \sum_{k\in \mathrm{s}}\sum_{l\in \mathrm{s}} \frac{\pi_{kl} - \pi_k\pi_l}{\pi_{kl}}\breve{u}_k\breve{u}_l.$$

Aprašytas Teiloro ištiesinimo metodas plačiai taikomas populiacijos dviejų sumų santykio, populiacijos dispersijos, kovariacijos ar regresijos koeficientų įvertinių apytikslėms dispersijoms apskaičiuoti.

1.5.2. Visrakčio metodas[2]

Pradžioje visrakčio metodas buvo naudotas įvertinių poslinkio minimizavimui begalinėje populiacijoje. 1958 metais Tukey [108] pasiūlė visrakčio metodą naudoti konstruojant populiacijos parametrų įvertinių dispersijų įvertinius. Jau po metų ši technika buvo pritaikyta baigtinėse populiacijose [22].

Nagrinėkime baigtinę populiaciją $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$ ir jos parametrą θ. Tegul $\widehat{\theta}$ – jo įvertinys. Tikslas – įvertinti parametro įvertinio dispersiją $Var(\widehat{\theta})$. Iš populiacijos $\mathcal{U}$ išrinkime fiksuoto dydžio n imtį s. Suskaidykime imtį s į A atsitiktinių grupių $\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_A$. Taigi $\mathrm{s} = \bigcup_{a=1}^{A} \mathrm{s}_a$. Paprastumo dėlei sakykime, kad grupių dydis $m = n/A$. Tarsime, kad kiekviena grupė $\mathrm{s}_j, j = 1, \ldots, A$, yra gauta paprasto atsitiktinio negrąžintinio ėmimo iš s metu. Kiekvienai gru-

[2] Angl. jackknife method.

pei apskaičiuojame $\widehat{\theta}_{(a)}$ – parametro θ įvertį pagal tą pačią funkcinę formą $\widehat{\theta}$ ir naudodamiesi informacija tų elementų, kurie liko po a-osios grupė išmetimo iš s. Skaičius $\widehat{\theta}_a = A\widehat{\theta} - (A-1)\widehat{\theta}_{(a)}$ vadinamas *a-ąja pseudoreikšme* ir naudojamas parametro θ visrakčio įvertiniui apibrėžti:

$$\widehat{\theta}_{JK} = \frac{1}{A}\sum_{a=1}^{A}\widehat{\theta}_a.$$

Pastarojo įvertinio dispersijos įvertiniu laikome

$$\widehat{Var}_{JK} = \frac{1}{A(A-1)}\sum_{a=1}^{A}(\widehat{\theta}_a - \widehat{\theta}_{JK})$$

ir vadiname *visrakčio dispersijos įvertiniu.*

Naudodami visrakčio dispersijos įvertinį galime įvertinti daugelio sudėtingų populiacijos parametrų įvertinių dispersijas, bet tai reikalauja daug kompiuterinio laiko.

Be aptartų metodų, įvertinių dispersijoms vertinti yra daugybė alternatyvų, kaip antai savirankos [90], atsitiktinių grupių metodai [85] ir kt.

1.6. Kalibruotieji populiacijos sumos įvertiniai. Atstumo funkcijos

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$, iš kurios išrinkta tikimybinė imtis s pagal imties planą $p(\mathrm{s})$. Tarsime, kad elementų priklausymo imčiai tikimybės π_k ir π_{kl} – teigiamos. Kaip ir anksčiau, y – tyrimo kintamasis, o $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ yra k-ojo populiacijos elemento J papildomų kintamųjų reikšmių vektorius. Tegul yra žinoma suma $\mathbf{t_x} = \sum_{k\in U}\mathbf{x}_k$ ir kiekvieno imties s elemento papildomų kintamųjų reikšmių vektorius $\mathbf{x}_k$. Suma $\mathbf{t_x}$ naudojama konstruojant specialius populiacijos sumos NPK įvertinius, kurie vadinami *kalibruotaisiais įvertiniais*. Kalibravimo idėją, modifikuojant Horvico ir Tompsono sumos įvertinio $\hat{t}_{y\pi} = \sum_{k\in\mathbf{s}} y_k/\pi_k = \sum_{k\in\mathbf{s}} d_k y_k$ svorius $d_k = 1/\pi_k$, pasiūlė Deville ir Särndal [21].

Kalibruotuoju populiacijos sumos t_y įvertiniu vadinsime įvertinį

$$\hat{t}_{yw} = \sum_{k\in\mathbf{s}} w_k y_k, \tag{1.10}$$

kuriame svoriai w_k:

1) tenkina lygtį
$$\sum_{k\in\mathbf{s}} w_k\mathbf{x}_k = \mathbf{t}_{\mathbf{x}},$$
kuri vadinama *kalibravimo lygtimi*;

2) minimizuoja atstumo funkciją
$$L(w_k, d_k, k\in\mathbf{s}) = \sum_{k\in\mathbf{s}} G_k(w_k, d_k)/q_k, \tag{1.11}$$

čia q_k – laisvai pasirenkami teigiami svoriai. Funkcija $G_k(w, d)$ tenkina šias sąlygas:

- kiekvienam fiksuotam $d > 0$, $G_k(w, d)$ yra neneigiama, diferencijuojama w atžvilgiu, griežtai išgaubta ir $G_k(d, d) = 0$;
- $g_k(w, d) = \frac{\partial G_k(w,d)}{\partial w}$ yra tolydi, griežtai didėjanti funkcija ir $g_k(d, d) = 0$.

Svoriai w_k vadinami *kalibruotaisiais svoriais*. Jei $G_k(w_k, d_k) = (w_k - d_k)^2/d_k$, tai gauname dažniausiai naudojamą atstumo funkciją
$$L_1(w_k, d_k, k\in\mathbf{s}) = \sum_{k\in\mathbf{s}} \frac{(w_k - d_k)^2}{d_k q_k}, \tag{1.12}$$

Minimizuojant pastarąją funkciją kai tenkinama $1)$ kalibruotųjų svorių sąlyga, gaunamas populiacijos sumos kalibruotasis įvertinys:
$$\hat{t}^{(1)}_{yw} = \hat{t}_{y\pi} + \Big(\sum_{k\in U}\mathbf{x}_k - \sum_{k\in\mathbf{s}} d_k\mathbf{x}_k\Big)' \Big(\sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k'\Big)^{-1} \sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k y_k. \tag{1.13}$$

Kaip jau buvo minėta, svoriai q_k yra laisvai pasirenkami. Jais naudojantis, įvertiniai yra modifikuojami. Atskiru atveju gaunami gerai žinomi NPK įvertiniai. Pavyzdžiui, tarkime, kad turime vieną papildomą kintamąjį $x : x_1, x_2, \ldots,$ x_N, o papildomus svorius parinkime taip: $q_k = \frac{1}{x_k}$. Tada (1.13) įvertinys įgaus išraišką
$$\hat{t}^{(1)}_{yw} = \frac{t_x}{\hat{t}_{x\pi}}\hat{t}_{y\pi}.$$
Tai yra gerai žinomas 1.2.1 skyrelyje apibrėžtas santykinis sumos įvertinys.

Kartais svoriai q_k yra tiesiog praleidžiami, priskiriant $\forall k \in \{1, 2, \ldots, N\}$, $q_k = 1$.

Kalibruoti sumos įvertiniai pritaikomi vertinant tokius populiacijos parametrus, kurie yra kelių populiacijoje apibrėžtų kintamųjų sumų funkcijos [73]. Tarkime, kad baigtinėje populiacijoje $\mathcal{U}$ turime M apibrėžtų tyrimo kintamųjų $y^{(1)}$,

$\ldots, y^{(M)}$, čia $y^{(j)}$ įgyjamos reikšmės yra $y_{j1}, \ldots, y_{jN}$. Tegul su kiekvienais k ir j, $k = 1, \ldots N$, $j = 1, \ldots M$, turime papildomų kintamųjų reikšmių vektorių $\mathbf{x}_k^{(j)} = (x_{1k}^{(j)}, \ldots, x_{m_j k}^{(j)})'$, čia m_j – tyrimo kintamojo $y^{(j)}$ papildomų kintamųjų skaičius. Visa tai turint ir naudojant kalibruotus sumos įvertinius yra sukonstruojamas populiacijos parametro

$$\theta = f(t_1, \ldots, t_M), \quad t_j = \sum_{k=1}^{N} y_{jk}, \tag{1.14}$$

įvertinys

$$\widehat{\theta} = f(\hat{t}_{1w^{(1)}}, \ldots, \hat{t}_{Mw^{(M)}}), \tag{1.15}$$

čia $\hat{t}_{jw^{(j)}} = \sum_{k \in \mathbf{s}} w_k^{(j)} y_{jk}$ – sumos t_j kalibruotas įvertinys. Pastebėsime, kad kiekvienai sumai t_j vertinti naudojamos skirtingos kalibruotų svorių sistemos. Taip pat būtina paminėti, kad kuo skaičius M yra didesnis, tuo sudėtingiau yra įvertinti įvertinio (1.15) dispersiją.

Konstruojant kalibruotus populiacijos sumos įvertinius vietoj tradicinės atstumo funkcijos $L_1(w_k, d_k, k \in \mathbf{s})$ gali būti naudojamos ir kitos. Kalibruotiesiems populiacijos sumos įvertiniams konstruoti Deville ir Särndal [21] pasiūlė tokias atstumo funkcijas:

$$\begin{aligned} L_1 &= \sum_{k \in \mathbf{s}} \frac{(w_k - d_k)^2}{d_k q_k}, & L_2 &= \sum_{k \in \mathbf{s}} \frac{w_k}{q_k} \log \frac{w_k}{d_k} - \frac{1}{q_k}(w_k - d_k), \\ L_3 &= \sum_{k \in \mathbf{s}} 2\frac{(\sqrt{w_k} - \sqrt{d_k})^2}{q_k}, & L_4 &= \sum_{k \in \mathbf{s}} -\frac{d_k}{q_k} \log \frac{w_k}{d_k} + \frac{1}{q_k}(w_k - d_k), \\ L_5 &= \sum_{k \in \mathbf{s}} \frac{(w_k - d_k)^2}{w_k q_k}. \end{aligned} \tag{1.16}$$

Jie taip pat nagrinėjo asimptotines kalibruotųjų įvertinių savybes ir gavo keletą svarbių, teorinių rezultatų, iš kurių svarbiausi yra šie:

1.1 teiginys. *Bet kurios atstumo funkcijos* $L(w_k, d_k, k \in \mathbf{s})$ *atveju kalibruotasis populiacijos sumos įvertinys* $\hat{t}_{yw}$*, apibrėžtas* (1.10) *lygybe, yra suderintas ir* $N^{-1}(\hat{t}_{yw} - \hat{t}_{y\pi}) = O_p(n^{-1/2})$.

1.2 teiginys. *Kalibruotasis populiacijos sumos įvertinys* $\hat{t}_{yw}$*, apibrėžtas* (1.10) *lygybe, yra asimptotiškai ekvivalentus apibendrintajam regresiniam įvertiniui ir to-*

dėl abu įvertiniai turi tą pačią asimptotinę dispersiją

$$Var_a(\hat{t}_{yw}) = \sum_{k=1}^{N}\sum_{l=1}^{N}(\pi_{kl} - \pi_k\pi_l)(d_k E_k^\star)(d_l E_l^\star), \tag{1.17}$$

čia

$$E_k^\star = y_k - \mathbf{x}_k' \Big(\sum_{i=1}^{N} q_i \mathbf{x}_i \mathbf{x}_i'\Big)^{-1} \sum_{i=1}^{N} q_i \mathbf{x}_i y_i.$$

Funkcijos

$$L_6 = \sum_{k\in\mathbf{s}} \frac{1}{q_k}\Big(\frac{w_k}{d_k} - 1\Big)^2, \qquad L_7 = \sum_{k\in\mathbf{s}} \frac{1}{q_k}\Big(\frac{\sqrt{w_k}}{\sqrt{d_k}} - 1\Big)^2 \tag{1.18}$$

naudojamos Plikuso straipsnyje [73].

Farrell ir Singh [31] kalibruotiesiems įvertiniams konstruoti pasiūlė ištisą klasę atstumo funkcijų:

$$L(w_k, d_k, k\in\mathbf{s}) = \frac{1}{2}\sum_{k\in\mathbf{s}} \frac{(w_k - d_k)^2}{d_k q_k^\star} \pm \frac{1}{2}\Phi^2 \sum_{k\in\mathbf{s}} \frac{(w_k)^2}{d_k q_k^\star}, \tag{1.19}$$

čia $q_k^\star$ yra laisvai pasirenkami svoriai, Φ – teigiamas dydis, priklausantis nuo keliamų reikalavimų įvertinio tikslumui. Didėjant Φ, kalibruoto įvertinio vidutinė kvadratinė paklaida mažėja, bet poslinkis didėja. Atskiru atveju, kai $\Phi = 0$ ir $q_k^\star = \frac{1}{2}q_k$, atstumo funkcija L sutampa su L_1.

Tame pačiame straipsnyje, be tradicinės kalibravimo lygties, siūlomos ir kitos (vieno papildomo kintamojo x atveju):

$$\begin{aligned} t_x \sum_{k\in\mathbf{s}} w_k &= N\hat{t}_{x\pi}\,, & \sum_{k\in\mathbf{s}} w_k x_k &= t_x(1+\Phi^2)^{-1}, \\ \sum_{k\in\mathbf{s}} w_k(x_k - t_x/N) &= 0, & t_x \sum_{k\in\mathbf{s}} w_k &= N(1+\Phi^2)\hat{t}_{x\pi}. \end{aligned}$$

Kalibruotieji įvertiniai, gauti minimizuojant (1.19) funkcijas ir naudojant vieną iš minėtų kalibravimo lygčių, aprėpia daug žinomų populiacijos sumos įvertinių. Pavyzdžiui, ieškant kalibruotųjų svorių, kurie tenkina kalibravimo lygtį $t_x \sum_{k\in\mathbf{s}} w_k = N\hat{t}_{x\pi}$ ir minimizuoja L (prieš antrą funkcijos dėmenį imant ženklą $+$), pasirinkus $q_k^\star = 1$, gaunamas gerai žinomas Murthy [68] populiacijos sumos įvertinys

$$\hat{t}_y = \hat{t}_{y\pi}(\hat{t}_{x\pi}/t_x).$$

Naudojant minėtas atstumo funkcijas ir kalibravimo lygtis, gaunama plati įvertinių klasė, į kurią įeina ne tik Murthy įvertinys, bet ir Searls [88], Singh ir Srivastava [93], nepaslinktasis Hartley ir Ross [40] bei kiti įvertiniai.

1.7. Kalibruotieji įvertiniai neatsakymų į apklausas atveju

Kaip jau buvo minėta, neatsakymai į apklausas sumažina tyrimo rezultatų tikslumą, tuo pačiu padidindami naudojamų įvertinių dispersiją. Literatūroje pateikiami du standartiniai metodai neatsakymų į apklausas neigiamam poveikiui sumažinti: *trūkstamų reikšmių įrašymo* ir *svertiniai*. Kartais šie abu metodai yra naudojami vienu metu [87]. Trumpai aptarsime svertinius metodus.

Tegul turime populiaciją $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$ ir joje apibrėžtą tyrimo kintamąjį $y\,:\,y_1, y_2, \ldots, y_N$. Kaip ir anksčiau, $\mathbf{s}$ yra pagal tam tikrą imties planą išrinkta tikimybinė imtis. Raide $\mathbf{r}$ pažymėkime atsakiusių į apklausą elementų aibę. Tikslas – įvertinti sumą $t_y = \sum_{k\in U} y_k$ neatsakymų į apklausas atveju ir turint papildomos informacijos.

Tarkime, kad turime stipriai koreliuotus su tyrimo kintamuoju y papildomus kintamuosius $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$. Pažymėkime $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ – populiacijos k-ojo elemento papildomų kintamųjų reikšmių vektorių. Yra išskiriami du „informacijos lygiai“, kurie vadinami Info-S ir Info-U [61], [62]:

1. Info-S: $\mathbf{x}_k$ yra žinoma visiems $k \in \mathbf{s}$.
2. Info-U: $\mathbf{x}_k$ yra žinoma visiems $k \in \mathbf{s}$ ir žinoma suma $\sum_{k\in U} \mathbf{x}_k$.

Tolesnė idėja yra tokia: Info-S atveju naudojant kalibravimo lygtį

$$\sum_{k\in\mathbf{r}} w_{sk}\mathbf{x}_k = \sum_{k\in\mathbf{s}} d_k\mathbf{x}_k,$$

o Info-U atveju – lygtį

$$\sum_{k\in\mathbf{r}} w_{Uk}\mathbf{x}_k = \sum_{k\in U} \mathbf{x}_k,$$

ir minimizuojant atstumo funkciją L_1, apibrėžtą (1.12) lygtimi, ieškoma kalibruotų svorių w_{sk} (Info-S), w_{Uk} (Info-U). Kartu sukonstruojami minėtus informacijos lygius atitinkantys kalibruotieji įvertiniai

$$\hat{t}_{ws} = \sum_{k \in \mathbf{r}} w_{sk} y_k,$$

$$\hat{t}_{wU} = \sum_{k \in \mathbf{r}} w_{Uk} y_k.$$

Atkreipsime dėmesį, kad įvertiniuose sumuojama tik pagal tuos imties elementus, kurie atsakė į apklausą.

Straipsnyje [62] Lundström ir Särndal nagrinėja įvertinių $\hat{t}_{ws}$, $\hat{t}_{wU}$ dispersijų vertinimo klausimą ir išveda sukonstruotų įvertinių poslinkio formules.

Pasiūlytas populiacijos sumos vertinimo būdas, esant neatsakymams į apklausas, vis dažniau taikomas daugelio šalių valstybinės statistikos tyrimams. Geriausi rezultatai yra pasiekiami, kai turima stipriai koreliuotų su tyrimo kintamaisiais papildomų kintamųjų. Kaip matyti iš kai kurių matematinio modeliavimo darbų [3], tiksliausi įvertiniai gaunami kombinuojant trūkstamų reikšmių įrašymo ir svertinius metodus.

1.8. Sudėtingesnių populiacijos parametrų vertinimas kalibruojant imties plano svorius

Be populiacijos sumos, statistiniuose tyrimuose svarbūs ir kiti parametrai: populiacijoje apibrėžtų dviejų tyrimo kintamųjų sumų santykis, populiacijos dispersija, kovariacija, pasiskirstymo funkcija ir kt. Yra nedaug darbų, kuriuose aptariamas šių parametrų vertinimas turint papildomos informacijos. Apžvelgsime kai kuriuos iš jų.

1.8.1. Kalibruotieji santykio įvertiniai baigtinėse populiacijose

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$ ir joje apibrėžtus du tyrimo kintamuosius $y : y_1, y_2, \ldots, y_N$ ir $z : z_1, z_2, \ldots, z_N$. Tegul t_y ir t_z reikš nežinomas populiacijos sumas

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

Tikslas – pasinaudojus kalibravimo idėja įvertinti dviejų sumų santykį

$$R = \frac{t_y}{t_z}.$$

Tarkime, kad $x_y : x_{y1}, \ldots, x_{yN}$ ir $x_z : x_{z1}, \ldots, x_{zN}$ yra atitinkamai tyrimo kintamųjų y ir z papildomi kintamieji.

Viena iš galimybių sukonstruoti kalibruotą dviejų sumų santykio įvertinį yra tokia [72]: nagrinėti santykio įvertinius, kurių pavidalas yra

$$\widehat{R}_w = \frac{\sum_{k \in \mathbf{s}} w_k y_k}{\sum_{k \in \mathbf{s}} w_k z_k},$$

čia svoriai w_k tenkina kalibravimo lygtį

$$\frac{\sum_{k \in \mathbf{s}} w_k x_{yk}}{\sum_{k \in \mathbf{s}} w_k x_{zk}} = \frac{\sum_{k=1}^{N} x_{yk}}{\sum_{k=1}^{N} x_{zk}}$$

ir minimizuoja pasirinktą atstumo funkciją. Šio įvertinio savybes empiriškai nagrinėjo Chadyšas ir Krapavickaitė [9]. Publikacijoje [55] kalibruotas santykio įvertinys lyginamas su tradiciniais santykio įvertiniais, kurių pavidalas yra

$$\widehat{R} = \frac{\hat{t}_y}{\hat{t}_z},$$

čia $\hat{t}_y$ bei $\hat{t}_z$ yra arba Horvico ir Tompsono, arba santykiniai, arba regresiniai sumų t_y ir t_z įvertiniai. Padaroma išvada, kad kalibruotas santykio įvertinys turi mažesnę vidutinę kvadratinę paklaidą nei tradiciniai įvertiniai.

Yra sukonstruoti populiacijos sumų santykio kalibruoti įvertiniai, naudojantys dvi svorių sistemas [73]:

$$\widehat{R}_w^{(1)} = \frac{\sum_{k \in \mathbf{s}} w_k^{(1)} y_k}{\sum_{k \in \mathbf{s}} w_k^{(2)} z_k},$$

čia svoriai $w_k^{(1)}$ ir $w_k^{(2)}$ tenkina kalibravimo lygtį

$$\frac{\sum_{k=1}^{N} x_{yk}}{\sum_{k=1}^{N} x_{zk}} = \frac{\sum_{k \in \mathbf{s}} w_k^{(1)} x_{yk}}{\sum_{k \in \mathbf{s}} w_k^{(2)} x_{zk}}$$

ir minimizuoja atstumo funkciją

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

Kombinuojant atstumo funkcijas $L_i,\ i = 1, \ldots, 7$, aprašytas (1.16), (1.18) lygybėmis, galima sudaryti ir daugiau funkcijų santykio įvertiniams su dviem svorių sistemomis konstruoti.

1.8.2. Horvico ir Tompsono sumos įvertinio dispersijos kalibruotasis įvertinys

Nagrinėkime baigtinėje populiacijoje $\mathcal{U}$ apibrėžto tyrimo kintamojo y, įgyjančio reikšmes $y_1, y_2, \ldots, y_N$, Horvico ir Tompsono sumos įvertinį

$$\hat{t}_{y\pi} = \sum_{k \in \mathbf{s}} d_k y_k, \tag{1.20}$$

čia $d_k = \frac{1}{\pi_k}$ yra imties plano svoriai. Tarkime, kad turime vieną papildomą kintamąjį x, įgyjantį reikšmes $x_1, x_2, \ldots, x_N$.

Yates ir Grundy [116] parodė, kad (1.20) įvertinio dispersiją galima užrašyti tokiu pavidalu:

$$Var_{YG}(\hat{t}_{y\pi}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2. \tag{1.21}$$

Dispersijos (1.21) įvertinys yra

$$\widehat{Var}_{YG}(\hat{t}_{y\pi}) = \frac{1}{2} \sum_{i \in \mathbf{s}} \sum_{j \in \mathbf{s}} d_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

čia $d_{ij} = (\pi_i \pi_j - \pi_{ij})/\pi_{ij}$.

Pasinaudoję Deville ir Särndal [21] svorių kalibravimo idėja, Singh, Horn, Chowdhury ir Yu [96] pasiūlė kalibruotąjį Horvico ir Tompsono sumos įvertinio dispersijos įvertinį

$$\widehat{Var}_{C}(\hat{t}_{y\pi}) = \frac{1}{2} \sum_{i \in \mathbf{s}} \sum_{j \in \mathbf{s}} w_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

čia svoriai w_k minimizuoja duotą atstumo funkciją ir tenkina šią kalibravimo lygtį:

$$\sum_{i \in \mathbf{s}} \sum_{j \in \mathbf{s}} w_{ij} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2. \tag{1.22}$$

Dešiniojoje (1.22) lygties pusėje esantis reiškinys – tai dviguba papildomo kintamojo sumos t_x Horvico ir Tompsono įvertinio dispersija.

1.8.3. Pasiskirstymo funkcijos ir kvantilių kalibruotieji įvertiniai

Ta pati Deville ir Särndal [21] kalibravimo idėja buvo pritaikyta baigtinėje populiacijoje $\mathcal{U}$ apibrėžto kintamojo $y\ :\ y_1, \ldots, y_N$ pasiskirstymo funkcijos $F_y(t) = N^{-1}\sum_{k\in U}\Delta(t-y_k)$ kalibruotam įvertiniui konstruoti [77]. Čia $\Delta(z)$ yra tokia funkcija, kad $\Delta(z) = 1$, jei $z \geq 0$, ir $\Delta(z) = 0$ – priešingu atveju. Paprastumo dėlei tarkime, kad turime tik vieną papildomą kintamąjį $x : x_1, \ldots, x_N$.

Pasiskirstymo funkcijos $F_y(t)$ kalibruotuoju įvertiniu vadiname įvertinį

$$\widehat{F}_{y,cal}(t) = \frac{1}{N}\sum_{k\in\mathbf{s}} w_k\Delta(t-y_k),$$

kuriame svoriai w_k minimizuoja atstumo funkciją $L(w_k, d_k, k \in \mathbf{s})$ (1.11) ir tenkina kalibravimo lygtį

$$\frac{1}{N}\sum_{k\in\mathbf{s}} w_k\Delta(t-x_{yk}) = F_{x_y}(t). \tag{1.23}$$

Konstruojant kalibruotuosius įvertinius reikalaujama, kad kalibruoti svoriai w_k kaip galima mažiau skirtųsi nuo imties plano svorių d_k. Tai pasiekiama minimizuojant tradicines atstumo funkcijas $L(w_k, d_k, k \in \mathbf{s})$. Harms [38] siūlo minimizuoti įvertinio $\widehat{F}_{y,cal}(t)$ ir empirinės pasiskirstymo funkcijos $F_{\mathbf{s},y}(t) = N^{-1}\sum_{k\in\mathbf{s}} d_k\Delta(t-y_k)$ atstumo funkciją:

$$Dist\big(\widehat{F}_{y,cal}, F_{\mathbf{s},y}\big) \to min_w.$$

Čia atstumo funkcija $Dist(F_1, F_2)$ galėtų būti tokia:

$$Dist(F_1, F_2) = \int_{-\infty}^{\infty}(F_1(t) - F_2(t))^2 dt.$$

Kintamojo y α lygmens kvantilį apibrėžėme (1.7) formule (žr. 1.4 poskyrį). Galimas ir kitas apibrėžimas: jei $F_y(t)$ yra kintamojo y pasiskirstymo funkcija, tai kintamojo y α lygmens kvantiliu vadiname

$$Q_\alpha = F^{-1}(\alpha).$$

Tada turint įvertintą pasiskirstymo funkciją, lieka vienas žingsnis iki kvantilių vertinimo. Atkreipsime dėmesį, kad tas vienas žingsnis gali būti labai ilgas, nes funkcija $\Delta(z)$, nuo kurios priklauso pasiskirstymo funkcija $F(t)$, yra laiptinė, dėl to $F(t)$ neturi atvirkštinės. Ši problema iš dalies išsprendžiama įvedus interpoliacinį kalibruotą pasiskirstymo funkcijos įvertinį [39].

Rueda ir bendraautorių [81] straipsnyje, žinant kiekvieno populiacijos elemento k papildomų kintamųjų $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$ reikšmių vektorių $\mathbf{x}_k = (x_{k1}, \ldots, x_{kJ})'$, yra sudaromas naujas kintamasis g, įgyjantis reikšmes $g_k = \widehat{\boldsymbol{\beta}}'\mathbf{x}_k, \quad k = 1, 2, \ldots, N$, čia dydis $\widehat{\boldsymbol{\beta}}$ yra svertinis daugialypės regresijos tarp y ir papildomų kintamųjų $x^{(j)},\ j = 1, 2, \ldots, J$, koeficientas:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{k \in \mathbf{s}} d_k q_k \mathbf{x}_k {\mathbf{x}_k}' \right)^{-1} \sum_{k \in \mathbf{s}} d_k q_k \mathbf{x}_k y_k.$$

Kintamasis g naudojamas kaip papildomas kintamasis kalibravimo lygtyje (1.23), apibrėžiant pasiskirstymo funkcijos įvertinio kalibruotus svorius.

1.9. Funkcinės formos metodas[3]

Estevao ir Särndal [27] straipsnyje sakoma, kad atstumo funkcijų minimizavimas nėra vienintelis būdas svoriams kalibruoti. Šiame poskyryje aptarsime minėtų autorių pasiūlytą naują svorių kalibravimo būdą, kuris vadinamas *funkcinės formos metodu*.

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$. Tegul $y : y_1, \ldots, y_N$ yra tyrimo kintamasis, o $\mathbf{x}_k = (x_{1k}, \ldots, x_{Jk})'$ – k-ojo populiacijos elemento J papildomų kintamųjų reikšmių vektorius. Vertinamas parametras yra populiacijos suma $t_y = \sum_{k=1}^{N} y_k$. Kiekvienam $k \in \mathbf{s}$ apibrėšime teigiamą konstantą q_k ir vektorių $\mathbf{z}_k = (z_{1k}, \ldots, z_{Jk})'$ taip, kad:

a) $dim(\mathbf{z}_k) = dim(\mathbf{x}_k) = J$ (vektorių $\mathbf{z_k}$ ir $\mathbf{x_k}$ dimensija būtų vienoda);

b) matricos $\sum_{k \in \mathbf{s}} q_k \mathbf{z}_k {\mathbf{x}_k}'$ determinantas būtų nelygus nuliui.

Vektorius $\mathbf{z}_k$ vadinamas *instrumentiniu vektoriumi* [86], [51].

Ieškosime kalibruotų svorių tokio pavidalo:

$$w_k = d_k + q_k \boldsymbol{\lambda}_{\mathbf{s}}' \mathbf{z}_k, \tag{1.24}$$

čia parametrai q_k ir $\mathbf{z}_k$ parenkami taip, kad tenkintų a) ir b) sąlygas. Vektorius $\boldsymbol{\lambda}_{\mathbf{s}}'$ netiesiogiai apibrėžiamas kalibravimo lygtimi

$$\sum_{k \in \mathbf{s}} w_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k, \tag{1.25}$$

kurią tenkina kalibruoti svoriai w_k. Funkcinė (1.24) forma yra naudojama dėl to,

[3] Angl. functional form approach.

kad tinkamai parinkdami parametrus q_k ir $\mathbf{z}_k$ galėtume kontroliuoti skirtumą tarp imties plano svorių d_k ir kalibruotų svorių w_k.

Tame pačiame straipsnyje yra įrodomas toks teiginys.

1.3 teiginys. *Svoriai $w_{k,kalf}$, kurie yra* (1.24) *pavidalo ir tenkina kalibravimo* (1.25) *lygtį, yra lygūs*

$$w_{k,kalf} = d_k + \Big(\sum_{l=1}^{N} \mathbf{x}_l - \sum_{l \in \mathbf{s}} d_l \mathbf{x}_l\Big)' \Big(\sum_{l \in \mathbf{s}} q_l \mathbf{z}_l \mathbf{x}_l'\Big)^{-1} q_k \mathbf{z}_k;$$

atitinkamas kalibruotasis populiacijos sumos įvertinys yra

$$\hat{t}_{calf} = \sum_{k \in \mathbf{s}} w_{k,kalf} y_k; \tag{1.26}$$

įvertinio $\hat{t}_{kalf}$ poslinkis yra $O(n^{-1})$ eilės.

Skirtingai parenkant parametrų q_k ir $\mathbf{z}_k$ reikšmes, gaunama ištisa kalibruotųjų įvertinių klasė, kurią žymėsime $KALF$. Atskiru atveju, kai $q_k = d_k$ ir $\mathbf{z}_k = \mathbf{x}_k$, gauname apibendrintąjį regresinį įvertinį.

Vektorių $\mathbf{z}_k$ komponentės dažniausiai yra papildomų kintamųjų reikšmių vektorių $\mathbf{x}_k$ funkcijos. Pavyzdžiui, kai $x_{jk} > 0, j = 1, \ldots, J$, vektorius $\mathbf{z}_k$ galėtų būti toks $\mathbf{z}_k = (x_{1k}^{p-1}, \ldots, x_{Jk}^{p-1})'$, čia p – teigiama konstanta.

Apibrėžus naują įvertinių klasę $KALF$, [27] straipsnyje nagrinėjami du klausimai:

1. Kaip įvertinių pagrindinės statistinės charakteristikos priklauso nuo parametrų q_k ir $\mathbf{z}_k$ parinkimo?

2. Apibendrintasis regresinis įvertinys priklauso kalibruotų įvertinių klasei $KALF$. Ar galima rasti sąlygas, kurioms esant apibendrintasis regresinis įvertinys turėtų mažiausią dispersiją klasėje $KALF$?

Atsakymą į pirmą klausimą iš dalies galima rasti ir Koto publikacijoje [52]. Čia rašoma, kad įvertinio $\hat{t}_{kalf}$ asimptotiškai minimali dispersija pasiekiama tada, kai instrumentiniai vektoriai parenkami pagal šią taisyklę:

$$\mathbf{z}_k = d_k^{-1} \sum_{l \in \mathbf{s}} (d_k d_l - d_{kl}) \mathbf{x}_l.$$

Särndal [84] nagrinėja bendresnę funkcinę formą:

$$w_k = d_k F(\boldsymbol{\lambda}_{\mathbf{s}}' \mathbf{z}_k),$$

čia F – bet kuri tolydi funkcija, pavyzdžiui, $F(u) = exp(u)$. Paėmus $F(u) = 1 + u$ ir $q_k = d_k, k = 1, \ldots, n$, gaunama (1.24) forma.

1.10. Modeliais pagrįsti kalibruotieji įvertiniai

Modeliais pagrįsto kalibravimo idėją vertinant populiacijos vidurkį pateikė Wu ir Sitter [115].

Tarkime, kad baigtinėje populiacijoje $\mathcal{U}$ apibrėžto kintamojo y reikšmės y_1, $\ldots$, y_N yra nepriklausomų atsitiktinių dydžių $Y_1, \ldots, Y_N$ realizacijos, o papildomų kintamųjų $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$ reikšmių vektoriai $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ – fiksuoti. Tegul tyrimo kintamąjį y su papildomais kintamaisiais $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$ sieja pusiau parametrinis modelis ξ_{sp}:

$$E_{\xi_{sp}}(y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\beta}), \qquad V_{\xi_{sp}}(y_i|\mathbf{x}_i) = v_i^2\sigma^2, \quad i = 1, \ldots, N, \tag{1.27}$$

čia $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ ir σ^2 yra nežinomi modelio parametrai, $\mu(\mathbf{x}, \boldsymbol{\beta})$ – žinoma funkcija, priklausanti nuo $\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(J)})$ ir $\boldsymbol{\beta}$, v_i – žinoma $\mathbf{x}_i$ funkcija.

Išrinkę imtį, imties elementų duomenis naudojame (1.27) modelio parametrams $\boldsymbol{\beta}$, σ^2 vertinti, o gautus įverčius $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$ – reikšmėms

$$\hat{y}_i = \mu(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}), \tag{1.28}$$

kurios yra vadinamos kintamojo y reikšmių y_i *prognozuojamomis reikšmėmis*, skaičiuoti. Apskaičiavę $\hat{y}_i$, $i = 1, 2, \ldots, N$, sakome, kad *suprognozavome* kintamojo y reikšmes.

Dabar jau galime apibrėžti modeliu ξ_{sp} pagrįstą kalibruotąjį baigtinės populiacijos vidurkio įvertinį:

$$\widehat{\mu}_{MC} = \frac{1}{N} \sum_{i \in \mathbf{s}} w_i y_i,$$

čia svoriai w_k, kaip ir anksčiau, minimizuoja tyrimui pasirinktą atstumo funkciją $L(w_k, d_k, k \in \mathbf{s})$, apibrėžtą (1.11) formule, ir tenkina šias kalibravimo lygtis:

$$\frac{1}{N} \sum_{i \in \mathbf{s}} w_i = 1, \qquad \sum_{i \in \mathbf{s}} w_i \mu(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{N} \mu(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}).$$

Vertinant populiacijos vidurkį įvertiniu $\widehat{\mu}_{MC}$, reikia žinoti visas papildomų kintamųjų populiacijos reikšmes. Skirtingai nei Deville ir Särndal [21] straipsnyje, apibrėžti kalibruoti svoriai w_k priklauso nuo tyrimo kintamojo y reikšmių. Vadinasi, atlikdami imčių tyrimą, kuriame turime kelis tyrimo kintamuosius, turėsime kiekvienam kintamajam atskirai suskaičiuoti kalibruotus svorius. Pastaroji problema išsprendžiama praplėtus neuroninių tinklų modeliu pagrįstų kalibruotų įvertinių klasę [65].

Wu ir Sitter [115] empiriškai lygino modeliu pagrįstą kalibruotą populiacijos vidurkio įvertinį su netiesiniu regresiniu įvertiniu, taikydami logtiesinį modelį. Išanalizavus eksperimento rezultatus, paaiškėjo, kad įvertinys $\widehat{\mu}_{MC}$ turi daug mažesnę dispersiją.

Turint vieną papildomą kintamąjį x, Farrell ir Singh [32] konstravo modeliu

$$E_{\xi_0}(y_i|x_i) = \beta x_i, \qquad V_{\xi_0}(y_i|x_i) = v(x_i)\sigma^2, \quad i = 1, \ldots, N, \tag{1.29}$$

pagrįstą kalibruotą Horvico ir Tompsono sumos įvertinio dispersijos $Var_{YG}(\hat{t}_{x\pi})$ (1.21) įvertinį tokio pavidalo:

$$\widehat{Var}_{wYG}(\hat{t}_{y\pi}) = \frac{1}{2}\sum_{i\in \mathbf{s}}\sum_{\substack{j\in \mathbf{s}\\ i\neq j}} w_{ij}\theta^\star_{ij}\Big(d_iy_i - d_jy_j\Big)^2. \tag{1.30}$$

Modelio (1.29) parametrai β ir σ^2 yra nežinomi, o v – žinoma x_i funkcija; $\theta^\star_{ij} = \pi_i\pi_j - \pi_{ij}$. Įvertinio (1.30) kalibruoti svoriai w_{ij} tenkina kalibravimo lygtį

$$\frac{1}{2}E_{\xi_0}\bigg(\sum_{i\in \mathbf{s}}\sum_{\substack{j\in \mathbf{s}\\ i\neq j}} w_{ij}\theta^\star_{ij}(d_iy_i - d_jy_j)^2\bigg) = \frac{1}{2}E_{\xi_0}\bigg(\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N} \theta^\star_{ij}(d_iy_i - d_jy_j)^2\bigg)$$

ir minimizuoja atstumo funkciją

$$L^\star = \sum_{i\in \mathbf{s}}\sum_{\substack{j\in \mathbf{s}\\ i\neq j}} (w_{ij} - d_{ij})^2/d_{ij}q_{ij}.$$

Wu [113] toliau tyrinėjo modeliu pagrįstus kalibruotus populiacijos vidurkio įvertinius, kalibravimui naudodamas lygtį

$$\sum_{i\in \mathbf{s}} w_iu(x_i) = \sum_{i=1}^{N} u(x_i),$$

čia $u(\cdot)$ – realiųjų skaičių aibėje apibrėžta ir realias reikšmes įgyjanti funkcija.

Buvo sprendžiamas uždavinys: su kokia funkcija $u(\cdot)$ įvertinys $\widehat{\mu}_{MC}$ turės mažiausią dispersiją. Uždavinys išspręstas ir gautas įvertinys pavadintas optimaliuoju kalibruotuoju populiacijos vidurkio įvertiniu. Toliau straipsnyje kalbama apie optimaliuosius pasiskirstymo funkcijos, populiacijos dispersijos ir populiacijos sumos apibendrintojo regresinio įvertinio dispersijos kalibruotus įvertinius.

1.11. Kalibruotieji įvertiniai skirtingų imties planų atveju

Didžioji dalis mokslo darbų apie svorių kalibravimą sudaryti iš dviejų dalių: teorinės ir eksperimentinės. Teorinėje dalyje kalbama apie kalibruotus įvertinius bet kokio imties plano atveju, o eksperimentinėje dalyje pasirenkamas konkretus imties planas ir atliekamas matematinis modeliavimas bei pateikiamos išvados. Šiame poskyryje pakalbėsime apie kalibruotus įvertinius tam tikrų konkrečių imties planų atveju.

1.11.1. Kalibruotieji populiacijos vidurkio įvertiniai, esant sluoksniniam ėmimui

Toks imties planas, kai suskaidžius populiaciją į kelias bendrų elementų neturinčias dalis, vadinamas *sluoksniais*, imtys renkamos iš kiekvieno sluoksnio atskirai, nepriklausomai nuo kitų, vadinamas *sluoksniniu ėmimu*. Suskaidykime baigtinę populiaciją $\mathcal{U}$, turinčią N elementų, į H sluoksnių. Skaičius N_h reikš h-ojo sluoksnio dydį, n_h – imties h-ajame sluoksnyje dydį. Taigi $\sum_{h=1}^{H} N_h = N$ ir $\sum_{h=1}^{H} n_h = n$. Tarkime, kad kiekviename sluoksnyje renkame paprastąją atsitiktinę be grąžinimo imtį. Simboliu W_h žymėsime dalį elementų, priklausančių h-ajam sluoksniui, t. y. $W_h = N_h/N$, o y_{hi} ir x_{hi} – atitinkamai i-ąją tyrimo kintamojo y bei papildomo kintamojo x reikšmę sluoksnyje U_h. Tracy, Singh ir Arnab [107] kalibruotąjį populiacijos vidurkio įvertinį sluoksninės imties atveju apibrėžia taip:

$$\widehat{\mu}_{st,cal} = \sum_{h=1}^{H} \Omega_h \bar{y}_h, \tag{1.31}$$

čia $\bar{y}_h = n_h^{-1} \sum_{i=1}^{n_h} y_{hi}$, o svoriai Ω_k minimizuoja atstumo funkciją

$$L_{sl} = \sum_{h=1}^{H} \frac{(\Omega_h - W_h)^2}{W_h Q_h}$$

ir tenkina šias dvi kalibravimo lygtis:

$$\sum_{h=1}^{H} \Omega_h \bar{x}_h = \sum_{h=1}^{H} W_h \mu_{xh},$$

$$\sum_{h=1}^{H} \Omega_h s_{xh}^2 = \sum_{h=1}^{H} W_h S_{xh}^2.$$

Čia $\bar{x}_h = n_h^{-1} \sum_{i=1}^{n_h} x_{hi}$ ir $s_{xh}^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$ yra atitinkamai papildomo kintamojo x imties vidurkis ir dispersija, o S_{xh}^2 – papildomo kintamojo x dispersija h-ajame sluoksnyje.

Jei kiekviename sluoksnyje papildomo kintamojo dispersija nėra žinoma, tai tada naudojama tik pirmoji kalibravimo lygtis.

Matematinio modeliavimo metu [107] buvo lyginta apibrėžto kalibruoto (1.31) įvertinio ir regresinio įvertinio vidutinė kvadratinė paklaida turint 4 skirtingas populiacijas. Paaiškėjo, kad šiose populiacijose kalibruoto įvertinio tikslumas yra nuo $4{,}46\%$ iki $13{,}08\%$ didesnis.

1.11.2. Kalibruotieji populiacijos vidurkio įvertiniai, esant dviejų fazių ėmimui

Nagrinėkime dviejų fazių ėmimą. Iš populiacijos $\mathcal{U}$, turinčios N elementų, išrinkime tikimybinę imtį $\mathbf{s}_1$ su žinomomis pirmos fazės elementų priklausymo imčiai tikimybėmis π_{1k}. Taigi pirmos fazės imties plano svoriai yra $d_{1k} = 1/\pi_{1k},\ k \in \mathbf{s}_1$. Tada iš $\mathbf{s}_1$ išrinkime poimtį $\mathbf{s}$ su žinomomis sąlyginėmis elementų priklausymo poimčiui tikimybėmis $\pi_{2k} = \pi_{k|\mathbf{s}_1}$. Antros fazės imties plano svoriai yra $d_{2k} = 1/\pi_{2k}$ ir dydis $d_k = d_{1k}d_{2k}$ – populiacijos k-ojo elemento imties plano svoris. Kaip ir anksčiau, tyrimo kintamąjį žymėsime y ir domėsimės šio kintamojo sumos t_y vertinimu.

Tarkime, kad turime du rinkinius skirtingo tipo papildomų kintamųjų, apibrėžtų populiacijoje $\mathcal{U}$: $x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(J_1)}$ ir $x_2^{(1)}, x_2^{(2)}, \ldots, x_2^{(J_2)}$. Populiacijos k-ojo elemento visų papildomų kintamųjų reikšmių vektorių žymėsime

$$\mathbf{x}_k = (x_{11k}, x_{12k}, \ldots, x_{1J_1k}, x_{21k}, x_{22k}, \ldots, x_{2J_2k})' = (\mathbf{x}'_{1k}, \mathbf{x}'_{2k})',$$

čia $\mathbf{x}_{1k} = (x_{11k}, x_{12k}, \ldots, x_{1J_1k})'$, $\mathbf{x}_{2k} = (x_{21k}, x_{22k}, \ldots, x_{2J_2k})'$. Estevao ir Särndal [26], [28], Hidiroglou ir Särndal [43] kalibruotų įvertinių

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

konstravimui dviejų fazių ėmime nagrinėjo tokius atvejus:

1) kai papildomų kintamųjų $x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(J_1)}$ reikšmių vektorių $\mathbf{x}_{1k}$ suma $\mathbf{t}_1 = \sum_{k \in U} \mathbf{x}_{1k}$ yra žinoma, o antrojo rinkinio papildomų kintamųjų $x_2^{(1)}, x_2^{(2)}, \ldots, x_2^{(J_2)}$ reikšmių vektorių $\mathbf{x}_{2k}$ suma $\mathbf{t}_2 = \sum_{k \in U} \mathbf{x}_{2k}$ – nežinoma;

2) kai papildomi kintamieji yra tokie, kad kiekvienam $k \in \mathbf{s}_1$ vektoriai $\mathbf{x}_{1k}$ ir $\mathbf{x}_{2k}$ yra žinomi;

3) kai papildomi kintamieji yra tokie, kad kiekvienam $k \in \mathbf{s}$ vektoriai $\mathbf{x}_{1k}$ ir $\mathbf{x}_{2k}$ yra žinomi.

Atsižvelgiant į tai, kaip naudojami papildomi kintamieji, išskiriami keli svorių kalibravimo būdai. Paminėsime du iš jų:

1. *Vieno žingsnio kalibravimas.* Imties plano svoriai d_k modifikuojami, kalibruotus svorius w_k apibrėžiant šia kalibravimo lygtimi:

$$\sum_{k \in \mathbf{s}} w_k (\mathbf{x}_{1k}, \mathbf{x}_{2k})' = \left(\sum_{k \in U} \mathbf{x}_{1k}, \sum_{k \in \mathbf{s}_1} d_{1k} \mathbf{x}_{2k} \right)'.$$

2. *Dviejų žingsnių kalibravimas.* Pirmajame žingsnyje svoriai d_{1k} pakeičiami pirmos fazės kalibruotais svoriais $w_{1k}, k \in \mathbf{s}_1$, tenkinančiais lygtį

$$\sum_{k \in \mathbf{s}_1} w_{1k} \mathbf{x}_{1k} = \sum_{k \in U} \mathbf{x}_{1k}.$$

Antrajame žingsnyje, naudojant jau žinomus svorius w_{1k}, apskaičiuojami galutiniai svoriai $w_k, k \in \mathbf{s}$, tenkinantys lygtį

$$\sum_{k \in \mathbf{s}} w_k (\mathbf{x}_{1k}, \mathbf{x}_{2k})' = \sum_{k \in \mathbf{s}_1} w_{1k} (\mathbf{x}_{1k}, \mathbf{x}_{2k})'.$$

Antruoju atveju naudojama daugiau žinomos papildomos informacijos, todėl galima tikėtis, kad įvertinys su svoriais, gautais po dviejų žingsnių kalibravimo, bus tikslesnis. Galbūt tai priklauso nuo ryšio tarp kintamųjų $\mathbf{x}_{1k}, \mathbf{x}_{2k}$ ir y_k? Tai nagrinėjo Estevao ir Särndal [28].

Wu ir Luan [114] toliau tęsė tyrimus kalibruotų ir modeliais pagrįstų įvertinių teorijos srityje ir apibrėžė optimalųjį kalibruotąjį populiacijos sumos įvertinį dviejų fazių ėmimo atveju.

1.12. Pirmojo skyriaus apibendrinimas. Uždavinių formulavimas

Apibendrinus šio skyriaus medžiagą, formuluojame šiuos uždavinius:

1. Taikant atstumo funkcijas L_i, $i = 2, 3, \ldots, 7$, išvesti kalibruotųjų sumos įvertinių svorių išraiškas.
2. Imties plano svorių kalibravimo idėją pritaikyti baigtinės populiacijos kovariacijos (dispersijos) kalibruotųjų įvertinių konstravimui. Pasiūlyti naujas kalibravimo lygtis.
3. Sukonstruoti baigtinės populiacijos kovariacijos įvertinius, kurie naudoja daugiau nei vieną svorių sistemą.
4. Spręsti sukonstruotų baigtinės populiacijos sumos ir kovariacijos įvertinių dispersijų vertinimo uždavinį.
5. Modeliuojant palyginti sukonstruotus įvertinius tarpusavyje ir su standartiniais atitinkamų parametrų įvertiniais.

2

Baigtinės populiacijos sluoksniavimas

Imčių metodų teorijoje parametrų įverčių tikslumas priklauso nuo daug dalykų: imties dydžio, naudojamo įvertinio, papildomos informacijos, imties plano ir kt. Dažnai tikslesni rezultatai gaunami naudojant sluoksninį ėmimą, apibrėžtą 1.11.1 skyrelyje. Šio tipo imties planą geriausiai taikyti tada, kai yra papildomos informacijos apie populiacijos struktūrą. Tinkamai suskaidžius populiaciją į sluoksnius, gaunami tikslesni parametrų įverčiai. Sluoksniuojant populiacijas iškyla daug klausimų. Kurio kintamojo atžvilgiu reikia atlikti sluoksniavimą? Koks turėtų būti optimalus sluoksnių skaičius? Kiek elementų rinkti iš kiekvieno sluoksnio? Kaip nustatyti sluoksnių ribas, kad parametrų įverčiai būtų kuo tikslesni? Kituose poskyriuose aptarsime paskutinį klausimą.

2.1. Klasikinis sluoksniavimo uždavinys

Nagrinėkime baigtinę populiaciją $\mathcal{U}= \{u_1, u_2, \ldots, u_N\}$. Tegul $y : y_1, \ldots, y_N$ yra populiacijoje $\mathcal{U}$ apibrėžtas tyrimo kintamasis, o $x : x_1, \ldots, x_N$ – papildomas kintamasis. Suskaidykime populiaciją $\mathcal{U}$ į H sluoksnių, čia H – fiksuotas žinomas skaičius. Pažymėkime h-ąjį sluoksnį simboliu U_h, o išrinktą iš populiacijos $\mathcal{U}$ sluoksninę imtį – raide $\mathrm{s}, \mathrm{s} \subset \mathcal{U}$. Iš kiekvieno sluoksnio U_h išrinkime paprastąją atsitiktinę imtį, kurią žymėsime s_h.

Taikant tinkamą sluoksniavimo strategiją, galima gauti tikslesnius parametrų įverčius, esant tai pačiai tyrimo kainai. Statistikų uždavinys yra nuspręsti, kaip pasirinkti geriausią sluoksniavimo algoritmą, kuriuo maksimizuojamas įvertinių tikslumas, t. y. minimizuojama įvertinių dispersija, vidutinė kvadratinė paklaida ar variacijos koeficientas.

Klasikinis sluoksniavimo uždavinys yra suformuluotas populiacijos vidurkio μ_y atveju minimizuojant jo įvertinio

$$\widehat{\mu}_y = \frac{1}{N}\sum_{h=1}^{H} N_h \bar{y}_h \tag{2.1}$$

dispersiją. Čia $\bar{y}_h$ yra h-ojo sluoksnio imties vidurkis, N_h – elementų skaičius h-ajame sluoksnyje, o $N_h\bar{y}_h$ – gerai žinomas h-ojo sluoksnio sumos Horvico ir Tompsono įvertinys.

Tarkime, kad imties dydis n yra iš anksto pasirinktas ir žinomas bei paskirstytas į sluoksnius pagal optimalųjį Neimano[1] paskirstymą [56].

Iš pradžių tarkime, kad kintamojo y reikšmės yra žinomos ir išrikiuotos didėjimo tvarka. Pažymėkime k_0 ir k_H atitinkamai mažiausią ir didžiausią y reikšmę. Uždavinys yra keliamas tokioms tarpinėms reikšmėms $k_1, \dots, k_{H-1}$ rasti, kurios minimizuoja vidurkio įvertinio dispersiją $Var(\widehat{\mu}_y)$. Tos tarpinės reikšmės vadinamos *sluoksnių ribomis*. Prielaida, kad tyrimo kintamasis y yra žinomas, nėra reali, todėl toliau sluoksniavimui naudosime papildomą kintamąjį x. Šis kintamasis turi būti stipriai koreliuotas su kintamuoju y. Principas išlieka tas pats: kintamojo x reikšmes išdėstome didėjimo tvarka ir ieškome sluoksnių ribų, kurios minimizuoja kintamojo x vidurkio įvertinio dispersiją $Var(\widehat{\mu}_x)$.

Dalenius [15], [17] parodė, kad tokios ribos egzistuoja ir tenkina šias lygtis:

$$\frac{(k_h-\mu_h)^2+S_h^2}{S_h} = \frac{(k_h-\mu_{h+1})^2+S_{h+1}^2}{S_{h+1}},\ h=1,\dots,H-1, \tag{2.2}$$

čia μ_h, S_h yra atitinkamai kintamojo x vidurkis ir standartinis nuokrypis h-ajame sluoksnyje. Iš viso turime $H-1$ lygtį. Maža to – ir μ_h, ir S_h priklauso nuo k_h. Taigi turime sudėtingas iteracines lygtis. Ieškant šių lygčių sprendinių nėra aišku, kaip pasirinkti pradinį k_h artinį; nėra žinoma, ar konverguos iteracinis procesas. Šios problemos paskatino statistikus atlikti tyrimus, kurių metu tiriamos iteracinių (2.2) lygčių sprendinio aproksimacijos. Tai būtų Dalenius ir Hodge [16], [17] *šaknies iš f* taisyklė, kurią taikant sluoksnių ribos parenkamos taip, kad kiekviename sluoksnyje suma, gauta sumuojant reikšmes $\sqrt{f(x_l)}$ (čia $f(t)$ – sluoksniavimo kintamojo x santykinių dažnių funkcija), būtų ta pati; Ekman [24] sluoksniavimo

[1] J. Neyman, 1894–1981, amerikiečių statistikas

taisyklė, kuri apibrėžia sluoksnių ribas taip, kad jos tenkintų lygtis

$$W_h(k_h - k_{h-1}) = C_H,\ h = 1, \dots, H,$$

čia W_h yra h-ojo sluoksnio svoris, C_H – konstanta, priklausanti nuo sluoksnių skaičiaus; Gunning ir Horgan [35] *geometrinė* sluoksniavimo taisyklė, paremta idėja suvienodinti sluoksnių variacijos koeficientus; Lavallée ir Hidiroglou [58] taisyklė, pagrįsta imties dydžio n minimizavimu, esant tam tikrai įvertinio paklaidai, ir kt.

2.2. Asimetrinių populiacijų sluoksniavimas

Žinomi papildomi kintamieji gali būti naudojami ne tik įvertinių konstravimui, bet ir imties plano sudarymo etape. Šiame poskyryje nagrinėsime efektyvaus asimetrinių populiacijų sluoksniavimo uždavinį. Vertinsime tyrimo kintamojo y vidurkį μ_y ir sluoksniavimo kintamuoju laikysime papildomą kintamąjį x.

Dydis

$$ac = \frac{1}{(N-1)S_y^3} \sum_{k=1}^{N} (y_k - \mu_y)^3$$

vadinamas populiacijos *asimetrijos koeficientu*. Jei $ac \neq 0$, tai populiacija vadinama *asimetrine*.

Nagrinėsime populiacijas, turinčias teigiamą asimetriją ($ac > 0$). Pasiūlysime *pataisytąjį geometrinį* sluoksniavimo metodą ir eksperimentiškai palyginsime su trimis kitais sluoksniavimo metodais: Dalenius ir Hodges [16], [17] *šaknies iš f* sluoksniavimo taisykle, Gunning ir Horgan [35] pasiūlytu *geometriniu* metodu bei *laipsnių* metodu, kurį įvedė Plikusas [71]. Šio poskyrio rezultatai paskelbti [A3] publikacijoje.

2.2.1. Sluoksniavimo metodai

Klasikinis sluoksniavimo uždavinys ir lygtys, kurios apibrėžia optimaliąsias sluoksnių ribas, pateikti 2.1 poskyryje. Išvardysime kelis šių lygčių sprendinį aproksimuojančius metodus.

I. *Šaknies iš f taisyklė.* Šiuo algoritmu randamos sluoksnių ribos, kurios yra labai arti (2.2) lygčių sprendinio.

Papildomo kintamojo x santykinių dažnių funkciją pažymėkime $f(t)$. Tarę, kad kiekviename sluoksnyje kintamasis x turi tolygųjį skirstinį, Dalenius ir Hodges

[14], [17] parodė, jog populiacijos vidurkio įvertinio minimali dispersija apytiksliai pasiekiama tada, kai h-ojo sluoksnio ribos $k_h^{(f)}$ parenkamos taip, kad sumos

$$\sum_{l \in U_h} \sqrt{f(x_l)}, \quad h = 1, 2, \ldots, H,$$

būtų apytiksliai lygios.

II. *Geometrinis metodas.* Vieną iš įdomesnių ir paprastų metodų pasiūlė Gunning ir Horgan [35]. Jų sluoksnių ribų konstravimo algoritmas pagrįstas pastebėjimu, kad jei sluoksnių ribos yra netoli optimaliųjų ribų, tai sluoksniavimo kintamojo x variacijos koeficientas kiekviename sluoksnyje yra toks pat:

$$\frac{S_1}{\mu_1} = \frac{S_2}{\mu_2} = \cdots = \frac{S_H}{\mu_H}.$$

Tarus, kad papildomas kintamasis x kiekviename sluoksnyje yra tolygiai pasiskirstęs, gaunamos apytikslės optimaliųjų sluoksnių ribų išraiškos:

$$k_h^{(g)} = k_0^{(g)} r^h, r = \left(\frac{k_H^{(g)}}{k_0^{(g)}} \right)^{1/H}, \quad h = 0, 1, \ldots, H.$$

Taigi sluoksnių ribos yra geometrinės progresijos nariai. Iš čia ir kilo pavadinimas *geometrinis* metodas. Šis metodas rekomenduotinas sluoksniuojant asimetrines populiacijas.

III. *Laipsnių metodas.* Paprastas ir efektyvus sluoksniavimo metodas siūlomas [71] straipsnyje. Čia sluoksnių ribos $k_h^{(p)}$ parenkamos taip, kad sumos

$$\sum_{l \in U_h} x_l^{\alpha^\star} \approx const, \quad h = 1, 2, \ldots, H,$$

būtų apytiksliai lygios. Tačiau šis metodas kol kas neturi teorinio pagrindimo ir jo privalumai gali būti parodyti tik atliekant matematinį modeliavimą. Daugybė eksperimentų rodo, kad, taikant laipsnių metodą, vidurkio įvertinio (2.1) dispersija yra minimizuojama, kai $0{,}5 < \alpha^\star < 0{,}7$. Net egzistuoja hipotezė, kad jei populiacija turi eksponentinį skirstinį, tai parametras $\alpha^\star$ priklauso nuo šio skirstinio parametro.

IV. *Pataisytasis geometrinis metodas*.

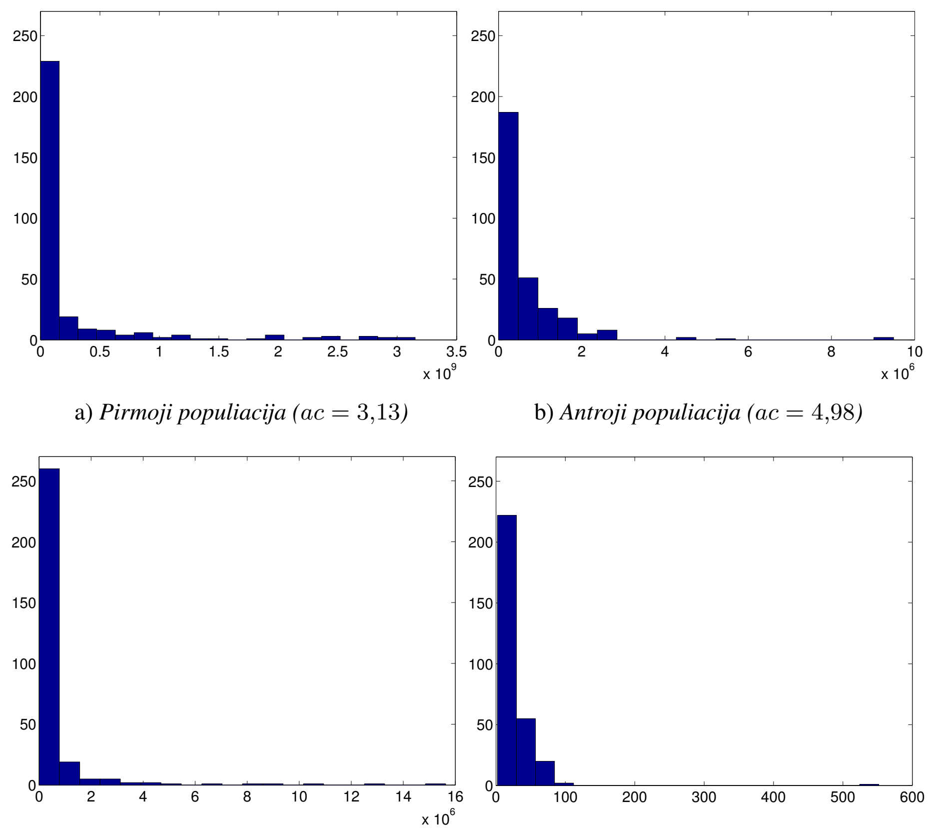
Praktikoje (pvz., atliekant darbo užmokesčio tyrimus) tiriamos populiacijos dažnai turi teigiamą asimetrijos koeficientą ir jų skirstinys yra artimas eksponentiniam skirstiniui. Modifikuosime geometrinį sluoksniavimo metodą tokio tipo populiacijų atveju. Padarę prielaidą, kad sluoksniavimo kintamojo skirstinys yra eksponentinis ir naudodami tą pačią Gunning ir Horgan [35] idėją suvienodinti variacijos koeficientus sluoksniuose, gauname iteracines lygtis, kurios apibrėžia sluoksnių ribas:

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

čia

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

Palyginkime aprašytuosius sluoksniavimo metodus taikydami matematinį modeliavimą.



a) *Pirmoji populiacija (*$ac = 3{,}13$*)* b) *Antroji populiacija (*$ac = 4{,}98$*)*

c) *Trečioji populiacija (*$ac = 5{,}95$*)* d) *Ketvirtoji populiacija (*$ac = 10{,}91$*)*

2.1 pav. *Tyrimo kintamojo* y *reikšmių histogramos*

2.1 lentelė. *Sluoksniavimo metodų palyginimas (sluoksnių skaičius $H = 5$, imties dydis $n = 50$)*

Populiacija	Sluoksniavimo metodas	cv		Sluoksniai 1	2	3	4	5
1	Cum$\sqrt{f}$	0,159 2	$k_h^{(f)}$	75 709	118 149,8	173 649,1	255 265,9	
			N_h	83	72	57	47	41
			n_h	9	8	8	11	14
	Geometrinis	0,178 5	$k_h^{(g)}$	57 826,9	91 532,2	144 883,2	229 330,8	
			N_h	51	62	79	56	52
			n_h	3	5	11	12	19
	Laipsnių	0,150 1	$k_h^{(p)}$	81 624	122 303	180 321	263 000	
			N_h	96	69	55	44	36
			n_h	12	8	9	10	11
	Pataisytasis geometrinis	0,175 5	$k_h^{(adj)}$	58 129,7	92 350,2	146 364,2	231 123,6	
			N_h	51	65	77	56	51
			n_h	3	6	10	13	18
2	Cum$\sqrt{f}$	0,163 5	$k_h^{(f)}$	181	354,7	571,9	832,4	
			N_h	78	67	66	51	38
			n_h	7	6	7	7	22
	Geometrinis	0,227 2	$k_h^{(g)}$	22,9	71,4	223,2	697,3	
			N_h	9	22	68	141	60
			n_h	0	0	4	24	22
	Laipsnių	0,159 9	$k_h^{(p)}$	261,2	464,3	649,1	872,3	
			N_h	117	62	48	40	33
			n_h	15	7	4	5	19
	Pataisytasis geometrinis	0,187 7	$k_h^{(adj)}$	25,6	88,1	288,4	849,7	
			N_h	11	25	89	140	35
			n_h	0	1	6	28	15
3	Cum$\sqrt{f}$	0,133 2	$k_h^{(f)}$	176 771,2	883 576	1 767 081,9	5 654 508,1	
			N_h	151	84	31	19	15
			n_h	4	8	3	11	24
	Geometrinis	0,228 8	$k_h^{(g)}$	842,4	10 138,1	122 006,8	1 468 292,4	
			N_h	1	30	95	136	38
			n_h	0	0	1	11	38
	Laipsnių	0,123 1	$k_h^{(p)}$	329 353	1 036 455	3 948 983	9 051 276	
			N_h	189	56	31	14	10
			n_h	8	6	11	11	14
	Pataisytasis geometrinis	0,180 8	$k_h^{(adj)}$	1 182	21 194,1	375 856,7	3 758 593,2	
			N_h	4	49	147	75	25
			n_h	0	0	4	13	32
4	Cum$\sqrt{f}$	0,173 0	$k_h^{(f)}$	6,8	15,6	24,4	39	
			N_h	79	100	61	37	23
			n_h	4	6	4	4	32

Populiacija	Sluoksniavimo metodas	cv		Sluoksniai 1	2	3	4	5
4	Geometrinis	0,069 3	$k_h^{(g)}$	3,1	9,7	30,2	94,1	
			N_h	54	39	169	37	1
			n_h	1	2	32	15	0
	Laipsnių	0,196 8	$k_h^{(p)}$	11	16	22	37	
			N_h	115	64	50	43	28
			n_h	9	2	2	4	32
	Pataisytasis geometrinis	0,064 5	$k_h^{(adj)}$	3,5	12,2	40	116,1	
			N_h	54	95	128	22	1
			n_h	2	9	30	9	0

2.2.2. Sluoksniavimo metodų palyginimas atliekant matematinį modeliavimą

Visus šiame poskyryje paminėtus sluoksniavimo metodus mes tarpusavyje lyginome atlikdami eksperimentą keturiose realiose populiacijose, turinčiose asimetrinius skirstinius (žr. 2.1 pav.) ir po 300 elementų. Imties dydis $n = 50$ buvo paskirstytas į penkis sluoksnius naudojantis Neimano optimaliuoju paskirstymu. Sluoksniavimui naudotas žinomas papildomas kintamasis x, o rezultatai pateikti tyrimo kintamajam y, kurio koreliacija su papildomu kintamuoju x yra stipri ($\rho \approx 0{,}9$).

Kiekvienu sluoksniavimo metodu buvo apskaičiuotos sluoksnių ribos ir kiekvienoje populiacijoje, išrinkus po $m = 1000$ imčių s_j, įvertintas vidurkio μ_y įvertinio variacijos koeficientas. Visų keturių nagrinėtų asimetrinių populiacijų matematinio modeliavimo rezultatai pateikti 2.1 lentelėje.

Kaip matyti iš rezultatų lentelės, daugumai asimetrinių populiacijų geriausias yra laipsnių metodas. Geometrinis metodas yra paprastas, bet daugeliu atveju populiacijos vidurkio μ_y įvertinio tikslumas yra mažiausias. Šią situaciją gerai iliustruoja rezultatai pirmos populiacijos atveju.

Antros ir trečios populiacijų asimetrijos koeficientai yra didesni, bet visų metodų efektyvumas išlieka beveik nepakitęs. Be to, dar labiau išryškėja skirtumai tarp laipsnių metodo ir kitų likusiųjų.

Vis dėlto turime paminėti, kad ypač asimetrinėms populiacijoms laipsnių metodas jau nėra geriausias. Tai gerai matyti ketvirtoje populiacijoje, turinčioje didžiausią asimetrijos koeficientą ac. Pataisytasis geometrinis metodas šiuo atveju yra tinkamesnis.

Mes taip pat atlikome matematinį modeliavimą populiacijose, turinčiose normalųjį skirstinį. Pastebėjome, kad tinkamiausias sluoksniavimo algoritmas tokio tipo populiacijose yra šaknies iš f taisyklė, tačiau efektyvumo prasme skirtumas tarp visų sluoksniavimo metodų yra minimalus.

2.3. Baigtinės populiacijos sluoksniavimas kelių tyrimo kintamųjų atveju

Dažnai statistiniuose tyrimuose turime daugiau nei vieną tyrimo kintamąjį, todėl šiuo atveju populiacijos sluoksniavimo teorija vienam tyrimo kintamajam negali būti pritaikyta. Šiame poskyryje aptarsime Rizvi, Gupta ir Bhargava [78] pasiūlytą sluoksniavimo techniką dviejų tyrimo kintamųjų atveju ir esant grąžintiniam ėmimui su tikimybėmis, proporcingomis papildomo kintamojo didumui.

Nagrinėkime baigtinę populiaciją $\mathcal{U}=\{u_1,u_2,\ldots,u_N\}$ ir joje apibrėžtus du tyrimo kintamuosius $y^{(1)}:y_{11},\ldots,y_{1N}$, $y^{(2)}:y_{21},\ldots,y_{2N}$. Tegul $x:x_1,\ldots,x_N$ yra bendras abiejų tyrimo kintamųjų papildomas kintamasis. Jis kartu bus ir sluoksniavimo kintamasis. Toliau nagrinėdami sluoksniavimo uždavinį simboliu y_{jhi} žymėsime tyrimo kintamojo $y^{(j)}$, $j=1,2$, i-ąją reikšmę sluoksnyje U_h, simboliu x_{hi} – papildomo kintamojo i-ąją reikšmę tame pačiame sluoksnyje U_h, o p_{hi} – i-ojo elemento iš h-ojo sluoksnio išrinkimo į imtį tikimybę. Ėmimo su tikimybėmis, proporcingomis papildomo kintamojo didumui, atveju ji yra lygi $p_{hi}=x_{hi}/t_{xh}$, čia t_{xh} – kintamojo x suma h-ajame sluoksnyje.

Kintamųjų $y^{(j)}$ vidurkių nepaslinktieji įvertiniai yra

$$\widehat{\mu}^{(j)}=\frac{1}{N}\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{n_h}\frac{y_{jhi}}{p_{hi}},\ \ j=1,2, \tag{2.3}$$

čia n_h – imties dydis h-ajame sluoksnyje.

Imties dydį n paskirsčius į sluoksnius pagal optimalųjį paskirstymo principą [14], įvertinių, apibrėžtų (2.3) lygybe, dispersijos ir kovariacija yra

$$Var(\widehat{\mu}^{(j)})=\frac{1}{nN^2}\sum_{h=1}^{H}\frac{1}{W_h}\left(\sum_{i=1}^{N_h}\frac{(y_{jhi})^2}{p_{hi}}-(t_{yh}^{(j)})^2\right),$$

$$Cov(\widehat{\mu}^{(1)},\widehat{\mu}^{(2)})=\frac{1}{nN^2}\sum_{h=1}^{H}\frac{1}{W_h}\left(\sum_{i=1}^{N_h}\frac{y_{1hi}y_{2hi}}{p_{hi}}-t_{yh}^{(1)}t_{yh}^{(2)}\right),$$

čia $t_{yh}^{(j)}$ – kintamojo $y^{(j)}$ suma h-ajame sluoksnyje.

Tarkime, kad kiekviename sluoksnyje tyrimo kintamųjų $y^{(j)},\ j=1,2$, priklausomybė nuo papildomo kintamojo x yra tiesinė ir regresijos tiesės eina per koordinačių pradžią, t. y. kintamieji susieti šiuo tiesiniu regresiniu modeliu:

$$y^{(j)}=\beta_{jh}x+e_j,\ j=1,2, \tag{2.4}$$

čia e_j yra tokios modelio paklaidos, kad $E(e_j|x) = 0,\ E(e_j e_j^{'}|xx^{'}) = 0$, jei $x \neq x^{'}$, ir $V(e_j|x) = \eta_j(t) > 0$ su visais t iš x reikšmių kitimo intervalo.

Optimaliąsias sluoksnių ribas pažymėkime $k_h = k_h(x),\ h = 1, \ldots, H-1$. Galiojant (2.4) modeliui ir minimizuojant funkciją

$$G_3 = \begin{vmatrix} Var(\widehat{\mu}^{(1)}) & Cov(\widehat{\mu}^{(1)}, \widehat{\mu}^{(2)}) \\ Cov(\widehat{\mu}^{(1)}, \widehat{\mu}^{(2)}) & Var(\widehat{\mu}^{(2)}) \end{vmatrix} \tag{2.5}$$

k_h atžvilgiu, Rizvi, Gupta ir Bhargava gautos lygtys vadinamos *minimaliosiomis lygtimis:*

$$\begin{aligned} &\left(\sum_{h=1}^{H} W_h \mu_{h\phi_1} \mu_{hx}\right)\{k_h \mu_{h\phi_2} + \mu_{hx}(\phi_2(k_h) - \mu_{h\phi_2})\} + \\ &+\left(\sum_{h=1}^{H} W_h \mu_{h\phi_2} \mu_{hx}\right)\{k_h \mu_{h\phi_1} + \mu_{hx}(\phi_1(k_h) - \mu_{h\phi_1})\} = \\ &= \left(\sum_{h=1}^{H} W_h \mu_{h\phi_1} \mu_{hx}\right)\{k_h \mu_{r\phi_2} + \mu_{rx}(\phi_2(k_h) - \mu_{r\phi_2})\} + \\ &+\left(\sum_{h=1}^{H} W_h \mu_{h\phi_2} \mu_{hx}\right)\{k_h \mu_{r\phi_1} + \mu_{rx}(\phi_1(k_h) - \mu_{r\phi_1})\}, \end{aligned} \tag{2.6}$$

čia $\phi_j(t) = \eta_j(t)/t; \mu_{h\phi_j}$ ir μ_{hx} yra atitinkamai $\phi_j(t)$ ir x vidurkiai sluoksnyje $U_h; r = h+1, h = 1, \ldots H-1$.

Sistemos (2.6) sprendinys $k_1, \ldots, k_{H-1}$ sudaro optimaliųjų sluoksnių ribų seką. Rasti sistemos tikslųjį sprendinį yra dar sudėtingiau nei aprašyto klasikinio sluoksniavimo uždavinio atveju. Straipsnyje [78] apytiksliai optimalioms sluoksnių riboms rasti pateikiama taisyklė įvedant funkciją

$$R_3(t) = (\mu_{\eta_1} \phi_2^{'}(t) + \mu_{\eta_2} \phi_1^{'}(t)) f(t),$$

čia $\mu_{\eta_j} = \int_a^b \eta_j(t) f(t) dt$, $f(t)$ – kintamojo x tankio funkcija, a ir b – mažiausia ir didžiausia x reikšmės. Tai vadinamoji *kubinės šaknies iš $R_3(t)$ taisyklė*. Ji teigia: jei funkcija $R_3(t)$ yra aprėžta ir egzistuoja dvi pirmosios jos išvestinės, tai sluoksnių ribas parinkus taip, kad

$$\int_{k_{h-1}}^{k_h} \sqrt[3]{R_3(t)} dt = \frac{1}{H} \int_a^b \sqrt[3]{R_3(t)} dt, \quad h = 1, \ldots, H,$$

gaunama apytiksliai optimalių sluoksnių ribų seka k_h.

Verma [109] taip pat nagrinėjo optimalaus sluoksniavimo uždavinį dviejų tyrimo kintamųjų atveju, kai imties dydžio n paskirstymas į sluoksnius yra vienodas ($n_1 = n_2 = \ldots = n_H$), o imties planas – paprastas atsitiktinis sluoksninis. Minimizuojant (2.5) funkciją, buvo gautos analogiškos (2.6) lygtims minimaliosios lygtys, apibrėžiančios optimaliąsias sluoksnių ribas.

2.4. Antrojo skyriaus apibendrinimas

1. Tobulindami sluoksniavimo metodus, mokslininkai turėtų atkreipti dėmesį į empirinį pastebėjimą, kad sluoksnių variacijos koeficientai yra apytiksliai vienodi, jei sluoksnių ribos artimos optimaliosioms.
2. Laipsnių metodas yra tinkamiausias vidutiniškai asimetrinėms populiacijoms sluoksniuoti. Tačiau jis kol kas neturi teorinio pagrindimo.
3. Taikant geometrinį sluoksniavimo metodą, populiacijos vidurkio įvertinio dispersija daugeliu atvejų yra didžiausia.
4. Pasiūlytas pataisytasis geometrinis asimetrinių populiacijų sluoksniavimo metodas.
5. Matematinio modeliavimo rezultatai rodo, kad ypač asimetrinėms populiacijoms (kai $ac > 10$) geriausias yra pataisytasis geometrinis sluoksniavimas.
6. Atliekant statistinius tyrimus, dažnai nagrinėjame daugiau nei vieną tyrimo kintamąjį. Į tai turėtų būti atsižvelgta kuriant modernius populiacijų sluoksniavimo metodus.

3

Kalibruotieji baigtinės populiacijos sumos ir kovariacijos įvertiniai

Siekiant gauti tikslesnius parametrų įverčius, baigtinių populiacijų statistikoje plačiai naudojami kalibruotieji įvertiniai, naudojantys papildomą informaciją. Šie įvertiniai taip pat naudojami oficialiojoje statistikoje, ypač socialiniuose tyrimuose. Kaip jau buvo minėta, kalibravimo idėją vertinant populiacijos sumą pasiūlė Deville ir Särndal [21]. Nesenai kalibravimo technika pradėta taikyti vertinant populiacijos sumą neatsakymų į apklausas atveju [61], [62]. Poskyryje 1.6 suformulavome kalibravimo uždavinį ir pateikėme keletą atstumo funkcijų, apibrėžtų (1.16) ir (1.18) lygybėmis, kurios yra naudojamos kalibruotiems svoriams rasti. Populiariausia ir daugiausia praktikoje taikoma yra atstumo funkcija $L_1(w_k, d_k, k \in \mathbf{s})$, aprašoma (1.12) formule. Ši funkcija yra pati paprasčiausia. Ja naudojantis randamas išreikštinis kalibravimo lygčių sprendinys tiek vertinant baigtinės populiacijos sumą, tiek dviejų sumų santykį [72]. Viena šios funkcijos neigiama savybė yra ta, kad kai kurioms populiacijoms galime gauti neigiamus kalibruotus svorius. Šiame skyriuje sukonstruosime kalibruotuosius populiacijos sumos [A2], [A6] ir kovariacijos įvertinius [A1], [A5], [A8]. Spręsime įvertinių dispersijų vertinimo uždavinį [A4], [A9], [A10]. Taip pat pateiksime matematinio modeliavimo rezultatus.

3.1. Kalibruotieji populiacijos sumos įvertiniai, esant skirtingoms atstumo funkcijoms

3.1.1. Populiacijos sumos įvertiniai ir jų apytikslės dispersijos

Nagrinėkime baigtinę populiaciją $\mathcal{U}=\{u_1, u_2, \ldots, u_N\}$, iš kurios išrinkta tikimybinė imtis s pagal imties planą $p(\mathbf{s})$. Tarkime, kad $y : y_1, y_2, \ldots, y_N$ yra populiacijoje $\mathcal{U}$ apibrėžtas tyrimo kintamasis, o $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$ – papildomi kintamieji, kurių reikšmių vektoriai $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})'$ yra žinomi imties s elementams. Vertinsime sumą

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

Pasinaudoję kalibruotojo sumos įvertinio (1.10) apibrėžimu ir atstumo funkcijomis L_i, apibrėžtomis (1.16) bei (1.18) lygybėmis, kitame teiginyje sukonstruosime septynis kalibruotus populiacijos sumos t_y įvertinius.

3.1 teiginys. *Svoriai $w_k^{(i)}$, $k \in \mathbf{s}$, $i = 1, 2, \ldots, 7$, kurie tenkina kalibravimo lygtį*

$$\sum_{k\in\mathbf{s}} w_k^{(i)} \mathbf{x}_k = \mathbf{t}_\mathbf{x} \tag{3.1}$$

ir minimizuoja atstumo funkciją L_i, apibrėžtą (1.16) *bei* (1.18) *lygybėmis, yra lygūs $w_k^{(i)} = d_k \nu_k^{(i)}$, čia*

$$\begin{aligned}
\nu_k^{(1)} &= 1 + q_k \Big(\sum_{l=1}^{N} \mathbf{x}_l' - \sum_{l\in\mathbf{s}} d_l \mathbf{x}_l'\Big)\Big(\sum_{l\in\mathbf{s}} d_l \mathbf{x}_l \mathbf{x}_l' q_l\Big)^{-1} \mathbf{x}_k, \\
\nu_k^{(2)} &= \exp\left(\mathbf{t}_\mathbf{x}'\left(\sum_{l\in\mathbf{s}} \frac{w_l^{(2)} q_l \mathbf{x}_l \mathbf{x}_l'}{-1 + \sqrt{2w_l^{(2)}/d_l - 1}}\right)^{-1} \mathbf{x}_k q_k\right), \\
\nu_k^{(3)} &= 4\left(2 - \mathbf{t}_\mathbf{x}'\left(\sum_{l\in\mathbf{s}} \frac{(w_l^{(3)})^{3/2} q_l \mathbf{x}_l \mathbf{x}_l'}{2(\sqrt{w_l^{(3)}} - \sqrt{d_l})}\right)^{-1} \mathbf{x}_k q_k\right)^{-2}, \\
\nu_k^{(4)} &= \left(1 - \mathbf{t}_\mathbf{x}'\left(\sum_{l\in\mathbf{s}} \frac{2w_l^{(4)} q_l \mathbf{x}_l \mathbf{x}_l'}{-1 + \sqrt{4w_l^{(4)}/d_l - 3}}\right)^{-1} \mathbf{x}_k q_k\right)^{-1},
\end{aligned} \tag{3.2}$$

$$\nu_k^{(5)} = \left(1 - \mathbf{t}_{\mathbf{x}}'\left(\sum_{l\in\mathbf{s}} \frac{3/4 w_l^{(5)} q_l \mathbf{x}_l \mathbf{x}_l^{'}}{-1/2 + \sqrt{(3w_l^{(5)})/(2d_l) - 5/4}}\right)^{-1} \mathbf{x}_k q_k\right)^{-1/2},$$

$$\nu_k^{(6)} = 1 + \Big(\sum_{l=1}^{N} \mathbf{x}_l^{'} - \sum_{l\in\mathbf{s}} d_l \mathbf{x}_l^{'}\Big)\Big(\sum_{l\in\mathbf{s}} d_l^2 \mathbf{x}_l \mathbf{x}_l^{'} q_l\Big)^{-1} d_k \mathbf{x}_k q_k,$$

$$\nu_k^{(7)} = \left(1 - \mathbf{t}_{\mathbf{x}}'\left(\sum_{l\in\mathbf{s}} \frac{(w_l^{(7)})^{3/2} d_l q_l \mathbf{x}_l \mathbf{x}_l^{'}}{\sqrt{w_l^{(7)}} - \sqrt{d_l}}\right)^{-1} d_k \mathbf{x}_k q_k\right)^{-2};$$

atitinkami kalibruotieji sumos įvertiniai yra

$$\hat{t}_{yw}^{(i)} = \sum_{k\in\mathbf{s}} d_k \nu_k^{(i)} y_k = \sum_{k\in\mathbf{s}} w_k^{(i)} y_k. \tag{3.3}$$

Įrodymas. *Formulę $\nu_k^{(1)}$ įrodė Deville ir Särndal [21]. Mes įrodysime teiginį tik atstumo funkcijos L_6 atveju, nes įrodymo eiga kitais atvejais yra analogiška.*

Užrašykime Lagranžo funkciją:

$$L^{(6)} = \sum_{k\in\mathbf{s}} \frac{1}{q_k}\left(\frac{w_k^{(6)}}{d_k} - 1\right)^2 - \boldsymbol{\lambda}^{(6)^{'}}\left(\sum_{k\in\mathbf{s}} w_k^{(6)} \mathbf{x}_k - \sum_{k=1}^{N} \mathbf{x}_k\right),$$

čia $\boldsymbol{\lambda}^{(6)}$ – Lagranžo daugiklių vektorius stulpelis. Dalinės išvestinės $\frac{\partial L^{(6)}}{\partial w_k^{(6)}}$ yra lygios nuliui, kai

$$\frac{2}{d_k q_k}\left(\frac{w_k^{(6)}}{d_k} - 1\right) - \boldsymbol{\lambda}^{(6)^{'}} \mathbf{x}_k = 0.$$

Iš čia gauname

$$w_k^{(6)} = d_k + \frac{1}{2} d_k^2 q_k \boldsymbol{\lambda}^{(6)^{'}} \mathbf{x}_k. \tag{3.4}$$

Dauginame (3.4) *lygtį iš dešinės pusės iš $\mathbf{x}_k^{'}$:*

$$w_k^{(6)} \mathbf{x}_k^{'} = d_k \mathbf{x}_k^{'} + \frac{1}{2} d_k^2 q_k \boldsymbol{\lambda}^{(6)^{'}} \mathbf{x}_k \mathbf{x}_k^{'}. \tag{3.5}$$

Vektoriaus $\boldsymbol{\lambda}^{(6)^{'}}$ išraiška gaunama susumavus (3.5) *lygtis ir pasinaudojus kalibra-*

vimo (3.1) *lygtimi:*

$$\boldsymbol{\lambda}^{(6)'} = 2\Big(\sum_{k=1}^{N} \mathbf{x}_k' - \sum_{k\in\mathbf{s}} d_k \mathbf{x}_k'\Big)\Big(\sum_{k\in\mathbf{s}} d_k^2 q_k \mathbf{x}_k \mathbf{x}_k'\Big)^{-1}.$$

Įrašę ją į (3.4) lygybę, įrodome teiginį.

□

Taikant sukonstruotus įvertinius praktikoje, yra naudinga turėti įvertinių dispersijų išraiškas arba mokėti šias dispersijas įvertinti. Nagrinėkime kalibruotą baigtinės populiacijos sumos įvertinį

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

gautą minimizuojant kurią nors funkciją

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

tenkinančią atstumo funkcijai keliamus reikalavimus (žr. 1.6 poskyrį). Remiantis Deville ir Särndal įrodytu 1.2 teiginiu, kalibruotojo įvertinio $\hat{t}_{yw}$ asimptotinė dispersija yra lygi

$$Var_a(\hat{t}_{yw}) = \sum_{k=1}^{N}\sum_{l=1}^{N}(\pi_{kl} - \pi_k\pi_l)(d_k E_k^\star)(d_l E_l^\star),$$

čia

$$E_k^\star = y_k - \mathbf{x}_k'\Big(\sum_{i=1}^{N} q_i \mathbf{x}_i \mathbf{x}_i'\Big)^{-1}\sum_{i=1}^{N} q_i \mathbf{x}_i y_i.$$

Įvertinio $\hat{t}_{yw}$ dispersijai vertinti Deville ir Särndal [21] pasiūlė šį dispersijos įvertinį:

$$\widehat{Var}(\hat{t}_{yw}) = \sum_{k\in\mathbf{s}}\sum_{l\in\mathbf{s}} \frac{\pi_{kl} - \pi_k\pi_l}{\pi_{kl}}(w_k e_k^\star)(w_l e_l^\star),$$

čia

$$e_k^\star = y_k - \mathbf{x}_k'\Big(\sum_{i\in\mathbf{s}} w_i q_i \mathbf{x}_i \mathbf{x}_i'\Big)^{-1}\sum_{i\in\mathbf{s}} w_i q_i \mathbf{x}_i y_i.$$

Kalibruotųjų sumos įvertinių apytiksles dispersijas galima apskaičiuoti naudojant Teiloro ištiesinimo metodą (žr. 1.5.1 skyrelį), bet ne visada šį metodą tiesiogiai galima taikyti, nes daugeliu atvejų kalibruotų svorių w_k negalima užrašyti išreikštiniu pavidalu. Konstruojant kalibruotus sumos įvertinius, išreikštinį svorių w_k pavidalą gavome tik atstumo funkcijų L_1 ir L_6, apibrėžtų atitinkamai (1.12) ir (1.18) lygybėmis, atvejais. Kitame teiginyje pateiksime šias funkcijas atitinkančių kalibruotųjų sumos įvertinių $\hat{t}_{yw}^{(1)}$, $\hat{t}_{yw}^{(6)}$ apytiksles dispersijas, apskaičiuotas taikant Teiloro ištiesinimo metodą.

3.2 teiginys. *Įvertinio $\hat{t}_{yw}^{(i)}$, i=1,6, apytikslė dispersija atstumo funkcijos L_i atveju yra*

$$AVar(\hat{t}_{yw}^{(i)}) = \sum_{k,l=1}^{N} (\pi_{kl} - \pi_k \pi_l) \frac{e_k^{(i)}}{\pi_k} \frac{e_l^{(i)}}{\pi_l}, \tag{3.6}$$

čia

$$e_k^{(1)} = y_k - \sum_{j=1}^{J} \Gamma_j^{(1)} x_{jk}, \quad e_k^{(6)} = y_k - \sum_{j=1}^{J} \Gamma_j^{(6)} x_{jk},$$

$$\Gamma_j^{(1)} = \boldsymbol{\tau}_j' \mathbf{A}_1^{-1} \mathbf{t}_{\mathbf{x}y}, \quad \Gamma_j^{(6)} = \boldsymbol{\tau}_j' \mathbf{A}_6^{-1} \mathbf{t}_{d\mathbf{x}y},$$

$$\boldsymbol{\tau}_j = (\tau_1, \tau_2, \ldots, \tau_J)', \quad \tau_i = 0, \ \textit{kai } i \neq j, \ \textit{o} \ \tau_j = 1,$$

$$\mathbf{A}_1 = \sum_{k=1}^{N} q_k \mathbf{x}_k \mathbf{x}_k', \quad \mathbf{A}_6 = \sum_{k=1}^{N} d_k q_k \mathbf{x}_k \mathbf{x}_k',$$

$$\mathbf{t}_{\mathbf{x}y} = \sum_{k=1}^{N} q_k \mathbf{x}_k y_k, \quad \mathbf{t}_{d\mathbf{x}y} = \sum_{k=1}^{N} d_k q_k \mathbf{x}_k y_k.$$

Įrodymas. *Teiginį įrodysime atstumo funkcijos L_6 atveju. Įrodymas atstumo funkcijos L_1 atveju yra analogiškas.*

Iš 3.1 teiginio seka, kad populiacijos sumos t_y įvertinys $\hat{t}_{yw}^{(6)}$ yra

$$\hat{t}_{yw}^{(6)} = \sum_{k \in \mathbf{s}} w_k^{(6)} y_k = \sum_{k \in \mathbf{s}} \bigg(d_k + \Big(\sum_{l=1}^{N} \mathbf{x}_l' - \sum_{l \in \mathbf{s}} d_l \mathbf{x}_l' \Big) \Big(\sum_{l \in \mathbf{s}} d_l^2 q_l \mathbf{x}_l \mathbf{x}_l' \Big)^{-1} d_k^2 q_k \mathbf{x}_k \bigg) y_k$$

$$= \sum_{k\in\mathbf{s}} d_k y_k + \Big(\sum_{k=1}^{N} \mathbf{x}'_k - \sum_{k\in\mathbf{s}} d_k \mathbf{x}'_k\Big)\Big(\sum_{k\in\mathbf{s}} d_k^2 q_k \mathbf{x}_k \mathbf{x}'_k\Big)^{-1} \sum_{k\in\mathbf{s}} d_k^2 q_k \mathbf{x}_k y_k.$$

Įvedus pažymėjimus

$$\widehat{\mathbf{A}}_6 = \sum_{k\in\mathbf{s}} d_k^2 q_k \mathbf{x}_k \mathbf{x}'_k, \quad \hat{\mathbf{t}}_{d\mathbf{x}y} = \sum_{k\in\mathbf{s}} d_k^2 q_k \mathbf{x}_k y_k,$$

įvertinys $\hat{t}_{yw}^{(6)}$ *įgyja pavidalą*

$$\hat{t}_{yw}^{(6)} = \hat{t}_{y\pi} + (\mathbf{t}'_{\mathbf{x}} - \hat{\mathbf{t}}'_{\mathbf{x}\pi})\widehat{\mathbf{A}}_6^{-1}\hat{\mathbf{t}}_{d\mathbf{x}y} = f(\hat{t}_{y\pi}, \hat{\mathbf{t}}_{\mathbf{x}\pi}, \widehat{\mathbf{A}}_6, \hat{\mathbf{t}}_{d\mathbf{x}y}).$$

Taigi $\hat{t}_{yw}^{(6)}$ *yra netiesinė funkcija, priklausanti nuo šių įvertinių:* $\hat{t}_{y\pi}$, $\hat{t}_{x^{(j)}\pi}$, $\hat{t}_{j0,\pi}$, $j = 1, 2, \ldots, J$, *bei* $\hat{t}_{ij,\pi}$, $i \le j = 1, 2, \ldots, J$, *kurie yra atitinkamai* J*-mačių vektorių* $\hat{\mathbf{t}}_{\mathbf{x}\pi}$, $\hat{\mathbf{t}}_{d\mathbf{x}y}$ *bei matricos* $\widehat{\mathbf{A}}_6$ *komponentės.*

Ištiesinsime funkciją $\hat{t}_{yw}^{(6)}$ *taikydami Teiloro ištiesinimo metodą. Šiam tikslui reikia šių dalinių išvestinių:*

$$\frac{\partial f}{\partial \hat{t}_{y\pi}} = 1,$$

$$\frac{\partial f}{\partial \hat{t}_{x^{(j)}\pi}} = -\boldsymbol{\tau}'_j \widehat{\mathbf{A}}_6^{-1}\hat{\mathbf{t}}_{d\mathbf{x}y}, \quad j = 1, 2, \ldots, J,$$

$$\frac{\partial f}{\partial \hat{t}_{ij,\pi}} = (\mathbf{t}'_{\mathbf{x}} - \hat{\mathbf{t}}'_{\mathbf{x}\pi})(-\widehat{\mathbf{A}}_6^{-1}\boldsymbol{\Lambda}_{ij}\widehat{\mathbf{A}}_6^{-1})\hat{\mathbf{t}}_{d\mathbf{x}y}, \quad i \le j = 1, 2, \ldots, J,$$

$$\frac{\partial f}{\partial \hat{t}_{j0,\pi}} = (\mathbf{t}'_{\mathbf{x}} - \hat{\mathbf{t}}'_{\mathbf{x}\pi})\widehat{\mathbf{A}}_6^{-1}\boldsymbol{\tau}_j, \quad j = 1, 2, \ldots, J.$$

čia $\boldsymbol{\tau}_j$ *–* J*-matis vektorius, kurio* j*-oji komponentė lygi* 1*, o visos kitos* $= 0$*;* $\boldsymbol{\Lambda}_{ij}$ *yra* $(J \times J)$*-osios eilės kvadratinė matrica, kurios elementai* (i, j) *ir* (j, i) *pozicijose lygūs* 1*, o visur kitur – nuliui.*

Pastebėję, kad

$$E\hat{t}_{y\pi} = t_y = \sum_{k=1}^{N} y_k, \quad E\hat{\mathbf{t}}_{\mathbf{x}\pi} = \mathbf{t}_{\mathbf{x}} = \sum_{k=1}^{N} \mathbf{x}_k,$$

$$E\widehat{\mathbf{A}}_6 = \mathbf{A}_6 = \sum_{k=1}^{N} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}, \quad E\hat{\mathbf{t}}_{d\mathbf{x}y} = \mathbf{t}_{d\mathbf{x}y} = \sum_{k=1}^{N} d_k q_k \mathbf{x}_k y_k,$$

tašką, kurio aplinkoje skleisime funkciją $\hat{t}_{yw}^{(6)}$ Teiloro eilute, pasirenkame taip:

$$\big(\hat{t}_{y\pi}, \hat{\mathbf{t}}_{\mathbf{x}\pi}, \widehat{\mathbf{A}}_6, \hat{\mathbf{t}}_{d\mathbf{x}y}\big) = \big(t_{y\pi}, \mathbf{t}_{\mathbf{x}\pi}, \mathbf{A}_6, \mathbf{t}_{d\mathbf{x}y}\big).$$

Ištiesiname funkciją $\hat{t}_{yw}^{(6)}$, imdami Teiloro skleidinio tiesinę dalį:

$$\hat{t}_{ywL}^{(6)} = t_y + 1(\hat{t}_{y\pi} - t_y) - \sum_{j=1}^{J} \Gamma_j^{(6)} (\hat{t}_{x^{(j)}\pi} - t_{x^{(j)}}) = \sum_{k\in\mathbf{s}} d_k y_k$$

$$- \sum_{j=1}^{J} \Gamma_j^{(6)} \Big(\sum_{k\in\mathbf{s}} d_k x_{jk} - t_{x^{(j)}} \Big) = \sum_{k\in\mathbf{s}} d_k e_k^{(6)} + D^{(6)},$$

čia

$$\Gamma_j^{(6)} = \boldsymbol{\tau}_j' \mathbf{A}_6^{-1} \mathbf{t}_{d\mathbf{x}y}, \quad e_k^{(6)} = y_k - \sum_{j=1}^{J} \Gamma_j^{(6)} x_{jk}, \quad D^{(6)} = \sum_{j=1}^{J} \Gamma_j^{(6)} t_{x^{(j)}}.$$

Įvertinio $\hat{t}_{yw}^{(6)}$ apytikslė dispersija yra

$$AVar\big(\hat{t}_{yw}^{(6)}\big) = Var\big(\hat{t}_{ywL}^{(6)}\big) = Var\Bigg(\sum_{k\in\mathbf{s}} d_k e_k^{(6)} + D^{(6)} \Bigg) = Var\Bigg(\sum_{k\in\mathbf{s}} d_k e_k^{(6)} \Bigg).$$

Teiginyje pateiktą sumos įvertinio $\hat{t}_{yw}^{(6)}$ apytikslės dispersijos išraišką gauname apskaičiavę kintamojo $e^{(6)}$ sumos Horvico ir Tompsono įvertinio dispersiją.

□

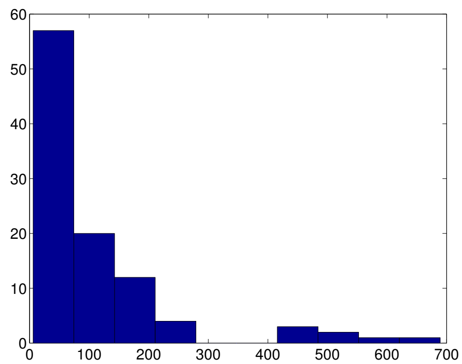
3.1 pastaba. Praktikoje galime taikyti šį dispersijos $Var(\hat{t}_{yw}^{(i)})$ įvertinį:

$$\widehat{Var}_L(\hat{t}_{yw}^{(i)}) = \sum_{k\in\mathbf{s}} \sum_{l\in\mathbf{s}} \Big(1 - \frac{\pi_k \pi_l}{\pi_{kl}}\Big) \frac{\hat{e}_k^{(i)}}{\pi_k} \frac{\hat{e}_l^{(i)}}{\pi_l}, \quad i = 1, 6.$$

Dydžiai $\hat{e}_k^{(1)}$, $\hat{e}_k^{(6)}$ yra apibrėžiami pakeičiant atitinkamose išraiškose $e_k^{(1)}$ ir $e_k^{(6)}$ esančius nežinomus parametrus $\mathbf{A}_1, \mathbf{t}_{\mathbf{x}y}$ ir $\mathbf{A}_6, \mathbf{t}_{d\mathbf{x}y}$ jų įverčiais $\widehat{\mathbf{A}}_1, \hat{\mathbf{t}}_{\mathbf{x}y}, \widehat{\mathbf{A}}_6, \hat{\mathbf{t}}_{d\mathbf{x}y}$.

3.1.2. Įvertinių tikslumo tyrimas atliekant matematinį modeliavimą

Buvo atliekamas eksperimentas. Pirmame bandyme nagrinėta dirbtinė $N = 100$ elementų populiacija, turinti skirstinį, artimą eksponentiniam (žr. 3.1 pav.). Populiacija suskaidyta į du sluoksnius: populiacijos elementai, atitinkantys mažesnes y reikšmes už kintamojo y vidurkį μ_y, sudaro pirmą sluoksnį, o likusieji elementai – antrą. Imties dydis $n = 20$. Naudojome dvimatį papildomų kintamųjų vektorių $\mathbf{x} = (1, x)'$ su reikšmėmis $\mathbf{x}_k = (1, x_k)'$, $k = 1, \ldots, N$. Eksperimento metu buvo lyginami sukonstruoti kalibruotieji populiacijos sumos įvertiniai $\hat{t}_{yw}^{(i)}$, $i = 1, 2, \ldots, 7$. Kiekvienam iš jų buvo išrenkama $m = 1000$ paprastųjų atsitiktinių sluoksninių imčių ir apskaičiuojami sumos įverčiai. Tyrėme, kaip įverčiai priklauso nuo skirtingų papildomų kintamųjų x, skirtingai koreliuojančių su tyrimo kintamuoju y. Rezultatų lentelėje 3.1 pateikiami gautųjų įverčių aritmetiniai vidurkiai.



3.1 pav. *Dirbtinė populiacija: tyrimo kintamojo y reikšmių histograma*

Turime paminėti, kad atstumo funkcijų L_2, L_3, L_4, L_5 ir L_7 atvejais ieškant kalibruotųjų svorių $w_k^{(i)}$, kuriuos apibrėžia iteracinės lygtys, gautoji sprendinio artinių seka kai kurioms išrinktoms imtims nekonverguoja. Todėl atlikdami matematinį modeliavimą, tokių imčių nenaudojome.

Kiekvienam populiacijos sumos kalibruotam įvertiniui apskaičiavome pagrindines empirines tikslumo charakteristikas (poslinkį $POSL$, dispersiją Var, vidutinę kvadratinę paklaidą VKP ir variacijos koeficientą cv), apibrėžiamas šiomis formulėmis:

$$\widehat{POSL}(\hat{t}_{yw}^{(i)}) = \frac{1}{m}\sum_{k=1}^{m} \hat{t}_{yw}^{(i)}(\mathbf{s}_k) - t_y,$$

3.1 lentelė. *Dirbtinė populiacija: pagrindinės kalibruotųjų sumos įvertinių tikslumo charakteristikos (tikroji sumos reikšmė $t = 10\,679$, imties dydis $n = 20$)*

Atstumo funkcija	Sumos įvertis	Dispersija $\times 10^{-6}$	$POSL$	VKP $\times 10^{-6}$	cv	Δ_{max}	Δ_{vid}
Koreliacijos koeficientas $\rho(y,x) = 0{,}8$							
L_1	10 713	2,342	33,58	2,343	0,143	0,086 0	0,026 9
L_2	10 726	2,339	46,56	2,341	0,143	0,058 8	0,019 0
L_3	10 725	2,339	46,28	2,341	0,143	0,058 7	0,018 9
L_4	10 727	2,340	47,67	2,342	0,143	0,059 9	0,019 2
L_5	10 729	2,341	49,91	2,344	0,143	0,061 8	0,019 6
L_6	10 712	2,340	33,41	2,341	0,143	0,080 1	0,028 5
L_7	10 725	2,337	46,35	2,339	0,143	0,055 9	0,021 4
Koreliacijos koeficientas $\rho(y,x) = 0{,}6$							
L_1	10 612	3,931	-66,76	3,935	0,187	0,081 0	0,026 7
L_2	10 629	3,931	-49,72	3,933	0,187	0,054 1	0,018 7
L_3	10 629	3,930	-49,96	3,933	0,187	0,054 0	0,018 6
L_4	10 630	3,931	-48,93	3,934	0,187	0,055 0	0,018 9
L_5	10 632	3,933	-47,24	3,935	0,187	0,056 6	0,019 3
L_6	10 611	3,928	-67,51	3,933	0,187	0,076 0	0,028 1
L_7	10 629	3,929	-50,01	3,931	0,186	0,050 9	0,021 1
Koreliacijos koeficientas $\rho(y,x) = 0{,}4$							
L_1	10 578	6,400	-101,46	6,411	0,239	0,081 6	0,027 1
L_2	10 604	6,430	-75,28	6,435	0,239	0,058 4	0,018 8
L_3	10 603	6,430	-75,52	6,435	0,239	0,058 3	0,018 7
L_4	10 605	6,431	-74,50	6,436	0,239	0,059 5	0,019 0
L_5	10 606	6,432	-72,81	6,438	0,239	0,061 6	0,019 5
L_6	10 576	6,396	-102,69	6,407	0,239	0,075 8	0,028 8
L_7	10 605	6,430	-73,53	6,435	0,239	0,055 9	0,021 0
Koreliacijos koeficientas $\rho(y,x) = 0{,}2$							
L_1	10 344	8,817	-335,06	8,930	0,287	0,091 9	0,026 3
L_2	10 376	8,896	-302,86	8,987	0,287	0,063 2	0,018 7
L_3	10 376	8,895	-303,08	8,987	0,287	0,063 1	0,018 6
L_4	10 377	8,896	-302,27	8,988	0,287	0,064 4	0,018 9
L_5	10 378	8,899	-300,87	8,990	0,287	0,066 7	0,019 4
L_6	10 344	8,818	-334,78	8,930	0,287	0,085 7	0,028 0
L_7	10 381	8,910	-298,46	9,000	0,288	0,060 3	0,020 8

$$\widehat{Var}(\hat{t}_{yw}^{(i)}) = \frac{1}{m}\sum_{k=1}^{m}\Big(\hat{t}_{yw}^{(i)}(\mathbf{s}_k) - \frac{1}{m}\sum_{l=1}^{m}\hat{t}_{yw}^{(i)}(\mathbf{s}_l)\Big)^2,$$

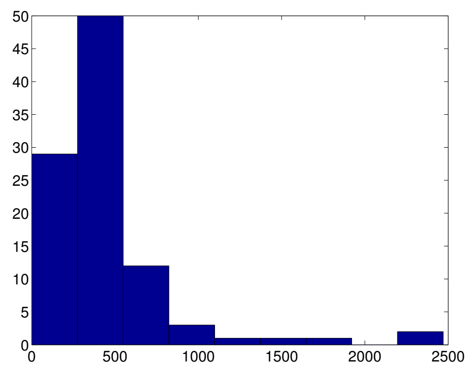
$$\widehat{VKP}(\hat{t}_{yw}^{(i)}) = \widehat{Var}(\hat{t}_{yw}^{(i)}) + (\widehat{POSL}(\hat{t}_{yw}^{(i)}))^2, \qquad \widehat{cv}(\hat{t}_{yw}^{(i)}) = \frac{\sqrt{\widehat{Var}(\hat{t}_{yw}^{(i)})}}{\frac{1}{m}\sum_{k=1}^{m}\hat{t}_{yw}^{(i)}(\mathbf{s}_k)},$$

čia m – išrinktų iš populiacijos $\mathcal{U}$ imčių skaičius, $\hat{t}_{yw}^{(i)}(\mathbf{s}_k)$ yra įvertinio $\hat{t}_{yw}^{(i)}$ reikšmė, apskaičiuota imčiai $\mathbf{s}_k$.

Taip pat buvo apskaičiuotos dvi svorių sklaidos charakteristikos:

$$\Delta_{max} = \max_{1 \leq k \leq n} |d_k - w_k| \quad \text{ir} \quad \Delta_{vid} = \frac{1}{n} \sum_{k=1}^{n} |d_k - w_k|. \tag{3.7}$$

Eksperimento rezultatai rodo, kad svorių w_k sklaida yra mažesnė naudojant atstumo funkcijas L_2, L_3, L_4, L_5 ir L_7. Jei lyginti svorių sklaidos charakteristikas atstumo funkcijų L_1 ir L_6 atvejais, tai pirmoji charakteristika yra mažesnė funkcijai L_6, o antroji – funkcijai L_1. Atstumo funkcijos L_3 ir L_7 garantuoja, kad kalibruoti svoriai bus teigiami. Tai yra teigiama savybė, nes neigiami svoriai neturi statistinės interpretacijos.



3.2 pav. *Reali populiacija: tyrimo kintamojo y reikšmių histograma*

Iš eksperimento rezultatų darome išvadą: kuo didesnė koreliacija tarp tyrimo ir papildomų kintamųjų, tuo mažesnę dispersiją ir vidutinę kvadratinę paklaidą turi įvertiniai.

Antro eksperimento metu naudotasi realia $N = 100$ elementų populiacija (žr. 3.2 pav.), kuri sudaryta taip pat iš dviejų sluoksnių. Populiacijos suskaidymui į sluoksnius taikyta ta pati taisyklė, kaip ir pirmos populiacijos atveju.

Iš pradžių kiekvienam įvertiniui buvo išrinkta 1000 paprastųjų sluoksninių $n = 20$ elementų imčių ir apskaičiuoti sumos įverčiai. Vėliau visa tai vėl buvo pakartota naujam 1000 paprastųjų sluoksninių $n = 40$ elementų imčių rinkiniui. Šiame eksperimente naudojome keturmatį papildomų kintamųjų vektorių $\mathbf{x} = (1, x^{(1)}, x^{(2)}, x^{(3)})'$ su reikšmėmis $\mathbf{x}_k = (1, x_{1k}, x_{2k}, x_{3k})'$, čia $x^{(1)}, x^{(2)}, x^{(3)}$ – kategoriniai kintamieji. Kaip ir pirmame bandyme, lyginome visus 7 kalibruotus sumos įvertinius. Kiekvienam iš jų įvertinome poslinkį, dispersiją, vidutinę

3.2 lentelė. *Reali populiacija: pagrindinės kalibruotųjų sumos įvertinių tikslumo charakteristikos (tikroji sumos reikšmė: $t = 45\,700$)*

Atstumo funkcija	Sumos įvertis	Dispersija $\times 10^{-7}$	$POSL$	VKP $\times 10^{-7}$	cv	Δ_{max}	Δ_{vid}
Imties dydis: $n = 20$							
L_1	45 451	6,905	-248,65	6,911	0,183	1,547 5	0,634 7
L_2	48 714	13,584	3 014,47	14,493	0,239	1,650 5	0,361 3
L_3	48 396	7,865	2 696,42	8,592	0,183	1,324 3	0,324 7
L_4	49 044	8,390	3 344,17	9,509	0,187	1,699 2	0,405 5
L_5	51 040	10,889	5 340,47	13,741	0,204	2,785 3	0,634 6
L_6	45 443	6,885	-256,86	6,892	0,183	1,827 9	0,636 4
L_7	48 384	7,753	2 684,58	8,474	0,182	1,622 3	0,324 2
Imties dydis: $n = 40$							
L_1	45 465	3,350	-234,52	3,356	0,127	0,689 1	0,286 4
L_2	48 218	3,725	2 517,95	4,359	0,127	0,611 2	0,153 0
L_3	48 141	3,703	2 441,44	4,299	0,126	0,588 3	0,147 5
L_4	48 710	3,900	3 009,90	4,806	0,128	0,736 1	0,179 0
L_5	50 335	5,128	4 635,03	7,276	0,142	1,180 6	0,271 7
L_6	45 461	3,338	-238,68	3,344	0,127	0,827 9	0,287 0
L_7	48 144	3,656	2 444,48	4,254	0,126	0,731 6	0,147 7

kvadratinę paklaidą, variacijos koeficientą ir apskaičiavome (3.7) svorių sklaidos charakteristikas (žr. 3.2 lentelę).

Kaip ir buvo tikėtasi, kuo didesnė imtis, tuo mažesnis įvertinių poslinkis, variacijos koeficientas, mažesnė dispersija, vidutinė kvadratinė paklaida ir svorių sklaida (žr. 3.2 lentelę).

Svarbu yra tai, kad naudojant keturmatį papildomų kintamųjų vektorių $(1, x^{(1)}, x^{(2)}, x^{(3)})$ kalibruotų svorių sklaida yra daug didesnė. Tai yra didelio poslinkio priežastis.

Iš gautų įvertinių tikslumo charakteristikų darome išvadą, kad įvertiniai $\hat{t}_{yw}^{(1)}$ ir $\hat{t}_{yw}^{(6)}$ yra šiek tiek tikslesni.

3.2. Baigtinės populiacijos kovariacijos kalibruotieji įvertiniai

Ankstesniame poskyryje nagrinėjome baigtinės populiacijos sumos kalibruotuosius įvertinius. Labiau sudėtingų parametrų vertinimas naudojant papildomą informaciją nėra plačiai nagrinėjamas literatūroje. Kaip jau buvo minėta, kalibruotus dviejų sumų santykio įvertinius nagrinėjo Plikusas [72], taip pat Krapavickaitė ir Plikusas [55]. Kalibruotus kvantilių įvertinius pasiūlė Harms ir Duchesne [39].

Sitter ir Wu [100] apibrėžė kalibruotus modeliais pagrįstus kovariacijos įvertinius.

Šiame poskyryje sukonstruosime keletą baigtinės populiacijos kovariacijos kalibruotųjų įvertinių, naudojančių vieną ar kelias svorių sistemas. Įvertiniams konstruoti naudojamos įvairios kalibravimo lygtys ir atstumo funkcijos. Tikslesni kovariacijos įvertiniai gali būti naudojami, pavyzdžiui, regresijos koeficientams vertinti.

Sukonstruotus įvertinius palyginsime empiriškai, atlikdami matematinį modeliavimą.

3.2.1. Pagrindiniai žymėjimai

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$, kurią sudaro N elementų. Paprastumo dėlei ir neprarandant bendrumo, galima laikyti $\mathcal{U} = \{1, 2, \ldots, N\}$. Tegul y ir z yra du tyrimo kintamieji, apibrėžti populiacijoje $\mathcal{U}$ ir atitinkamai įgyjantys reikšmes $\{y_1, y_2, \ldots, y_N\}$ ir $\{z_1, z_2, \ldots, z_N\}$. Kintamųjų y ir z reikšmės nėra žinomos. Domėsimės baigtinės populiacijos kovariacijos

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

vertinimu. Čia

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

Pažymėkime raide $\mathrm{s} \subset \mathcal{U}$ tikimybinę imtį, išrinktą iš populiacijos $\mathcal{U}$, o simboliu π_k – populiacijos k-ojo elemento priklausymo imčiai s tikimybę. Tegul $d_k = 1/\pi_k$ yra k-ojo elemento imties plano svoris.

Nagrinėkime du baigtinės populiacijos kovariacijos standartinius įvertinius:

$$\begin{aligned}
\widehat{Cov}_1(y,z) &= \frac{1}{N-1}\sum_{k\in\mathbf{s}} d_k\Big(y_k - \frac{1}{N}\sum_{l\in\mathbf{s}} d_l y_l\Big)\Big(z_k - \frac{1}{N}\sum_{l\in\mathbf{s}} d_l z_l\Big),\\
\widehat{Cov}_2(y,z) &= \frac{1}{N}\sum_{k\in\mathbf{s}} \frac{y_k z_k}{\pi_k} - \frac{1}{N(N-1)}\sum_{k,l\in s,\ k\neq l} \frac{y_k z_l}{\pi_{kl}},
\end{aligned} \tag{3.8}$$

čia π_{kl} žymi dviejų elementų poros (k, l) priklausymo imčiai tikimybę.

Įvertinius $\widehat{Cov}_1$ ir $\widehat{Cov}_2$ nagrinėjo Särndal, Swensson ir Wretman [85]. Mes juos naudosime palyginimui su kalibruotaisiais įvertiniais, kuriuos apibrėšime kitame skyrelyje.

3.2.2. Populiacijos kovariacijos įvertinių kalibruotieji svoriai

Tarkime, kad turime du papildomus kintamuosius a ir b, įgyjančius atitinkamai reikšmes $\{a_1, a_2, \ldots, a_N\}$ ir $\{b_1, b_2, \ldots, b_N\}$. Vadinasi, kiekvienam populiacijos elementui k turime papildomų kintamųjų reikšmių vektorių $\mathbf{a}_k = (a_k, b_k)'$. Tegul a yra papildomas kintamasis tyrimo kintamojo y, o b – kintamojo z. Žinomą papildomų kintamųjų kovariaciją pažymėkime $Cov(a, b)$. Naudodami žinomus papildomus kintamuosius, sukonstruosime naujus kalibruotus kovariacijos $Cov(y, z)$ įvertinius. Remdamiesi gautais kalibruotaisiais sumos įvertiniais, manome, kad ir kalibruotieji kovariacijos įvertiniai bus tikslesni už minėtus standartinius įvertinius $\widehat{Cov}_i(y, z)$, $i = 1, 2$, jei tik tyrimo ir atitinkami papildomi kintamieji bus gerai koreliuoti.

Nagrinėkime kalibruotuosius kovariacijos įvertinius $\widehat{Cov}_w(y, z)$, kurių pavidalas yra

$$\widehat{Cov}_w(y, z) = \frac{1}{N-1} \sum_{k \in \mathbf{s}} w_k (y_k - \widehat{\mu}_{yw})(z_k - \widehat{\mu}_{zw}), \tag{3.9}$$

čia

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

Kalibruotieji svoriai w_k yra apibrėžiami panašiai, kaip ir konstruojant kalibruotus sumos įvertinius:

a) svoriai w_k tenkina tam tikras kalibravimo lygtis;

b) atstumas tarp svorių d_k ir w_k yra minimalus. Tai pasiekiama minimizuojant vieną iš minėtų atstumo funkcijų $L_i(w_k, d_k, k \in \mathbf{s}), i = 1, \ldots, 7$.

Mes nagrinėjome tris kalibruotų baigtinės populiacijos kovariacijos įvertinių grupes, kurios apibrėžiamos naudojantis trimis skirtingomis kalibravimo lygtimis. Kiekvienai kalibravimo lygčiai naudotos keturios atstumo funkcijos (sąlyga b)).

I. Netiesinis kalibravimas. Tegul kalibravimo lygtis yra

$$\frac{1}{N-1} \sum_{k \in \mathbf{s}} w_k (a_k - \widehat{\mu}_{aw})(b_k - \widehat{\mu}_{bw}) = Cov(a, b), \tag{3.10}$$

čia

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

Pasiūlyta (3.10) kalibravimo lygtis yra netiesinė svorių w_k atžvilgiu, todėl šį atvejį pavadinome *netiesiniu kalibravimu*.

II. Tiesinis kalibravimas. Antroji kalibravimo lygtis yra

$$\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(a_k-\mu_a)(b_k-\mu_b)=Cov(a,b), \tag{3.11}$$

čia

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

Šį atvejį pavadinome *tiesiniu kalibravimu* todėl, kad kalibravimo lygtis, rodanti, kad mes iš tikrųjų kalibruojame kintamojo $(a-\mu_a)(b-\mu_b)$ sumą, yra tiesinė svorių w_k atžvilgiu.

III. Sumų kalibravimas. Trečio tipo kalibruotųjų svorių sistema apibrėžiama kalibruojant papildomų kintamųjų sumas. Kalibravimo lygtys yra tokios:

$$\sum_{k\in\mathbf{s}} w_k a_k=\sum_{k=1}^{N}a_k, \quad \sum_{k\in\mathbf{s}} w_k b_k=\sum_{k=1}^{N}b_k. \tag{3.12}$$

Šis atvejis remiasi dabartinių tyrimų praktika, kai naudojami kalibruoti įvertiniai ir tie patys svoriai taikomi visiems reikalingiems parametrams vertinti.

Suformuluosime teiginį ir įrodysime iteracines lygtis kalibruotiems svoriams w_k rasti. Tuo atveju, kai egzistuoja šių lygčių sprendinys, svoriai, gauti iteracinės procedūros metu, yra kalibruoti.

3.3 teiginys. *Svoriai* $w_k=w_k^{(i)}$, $k\in\mathbf{s}$, $i=1,3,6,7$, *kurie tenkina* (3.10) *lygtį ir minimizuoja atstumo funkciją* L_i, *apibrėžtą* (1.16) *bei* (1.18) *lygybėmis, tenkina lygtį* $w_k^{(i)}=d_k g_k^{(i)}$. *Čia:*

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

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

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

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

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

$\lambda^{(3)}$ *yra tinkamai parinkta lygties* $\alpha_2\lambda^2 + \alpha_1\lambda + \alpha_0 = 0$ *šaknis, čia*

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

$\lambda^{(6)} = \widehat{A}\Big(\sum_{k\in\mathbf{s}} d_k^2 q_k a_k b_k e_k\Big)^{-1}$; $\lambda^{(7)}$ *yra tinkamai parinkta lygties* $\beta_2\lambda^2+\beta_1\lambda+\beta_0 = 0$ *šaknis, čia*

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

Įrodymas. *Įrodymui paimkime atstumo funkciją* L_1 *ir apibrėžkime Lagranžo funkciją*

$$\Lambda_1 = \sum_{k\in\mathbf{s}} \frac{(w_k - d_k)^2}{d_k q_k} - \lambda\Big(\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(a_k - \widehat{\mu}_{aw})(b_k - \widehat{\mu}_{bw}) - Cov(a,b)\Big).$$

Spręsdami lygtį

$$\frac{\partial\Lambda_1}{\partial w_k} = 0, \quad k \in \mathbf{s},$$

gauname

$$w_k = d_k(1 + \frac{1}{2}\lambda q_k e_k). \tag{3.13}$$

Tada padauginę (3.13) lygtį iš $a_k b_k$*, susumavę pagal imties elementus ir panaudoję (3.10) kalibravimo lygtį, gauname* λ *išraišką. Įrašę ją į (3.13) lygybę, gauname iteracines lygtis svoriams* $w_k^{(1)}$ *rasti.*

Atstumo funkcijos L_3 *atveju sukonstruojame Lagranžo funkciją*

$$\Lambda_3 = \sum_{k\in\mathbf{s}} 2\frac{(\sqrt{w_k} - \sqrt{d_k})^2}{q_k} - \lambda\Big(\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k(a_k - \widehat{\mu}_{aw})(b_k - \widehat{\mu}_{bw}) - Cov(a,b)\Big).$$

Funkcijos Λ_3 *dalinės išvestinės yra lygios nuliui, kai*

$$w_k = d_k\big(\frac{1}{2}\lambda q_k e_k - 1\big)^{-2} \tag{3.14}$$

arba

$$\frac{1}{4}q_k^2 e_k^2\lambda^2 - q_k e_k\lambda - \frac{d_k}{w_k} + 1 = 0. \tag{3.15}$$

Padauginę (3.15) lygtį iš $w_k a_k b_k$*, susumavę pagal imties elementų aibę* s *ir panau-*

dotę (3.10) kalibravimo lygtį, gauname lygtį $\alpha_2\lambda^2 + \alpha_1\lambda + \alpha_0 = 0$*, kuri apibrėžia* $\lambda^{(3)}$*. Kalibruotų svorių* $w_k^{(3)}$ *išraiška išplaukia iš (3.14) lygybės.*

Įrodymo eiga atstumo funkcijų L_6 *ir* L_7 *atveju yra analogiška.*

□

Įveskime keletą papildomų žymėjimų: $c_k = (a_k - \mu_a)(b_k - \mu_b)$,

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

3.4 teiginys. *Svoriai* $w_k = w_k^{(i)}$*,* $k \in \mathbf{s}$*,* $i = 1, 3, 6, 7$*, kurie tenkina* (3.11) *lygtį ir minimizuoja atstumo funkciją* L_i*, apibrėžtą* (1.16) *bei* (1.18) *lygybėmis, tenkina lygtį* $w_k^{(i)} = d_k f_k^{(i)}$*. Čia:*

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

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

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

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

Įrodymas yra panašus kaip ir 3.3 teiginio.

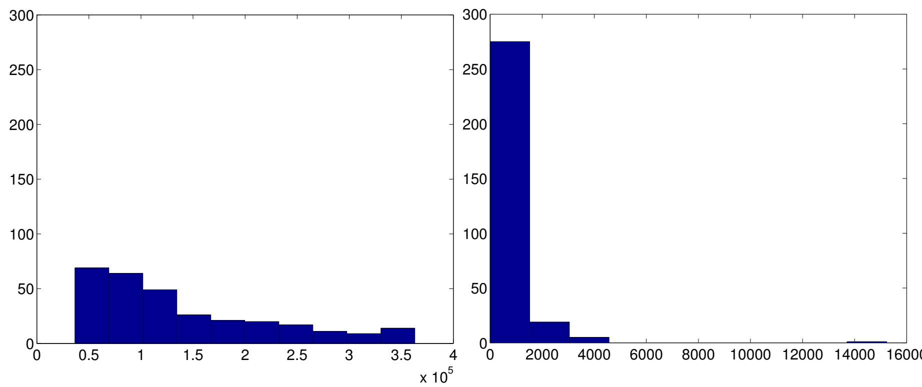
Sumų kalibravimo atveju (3.9) kovariacijos įvertinio svoriai w_k tiesiogiai randami iš 3.1 teiginio.

Kalibruotųjų svorių w_k išraišką kitų atstumo funkcijų atveju galima rasti naudojantis ta pačia technika.

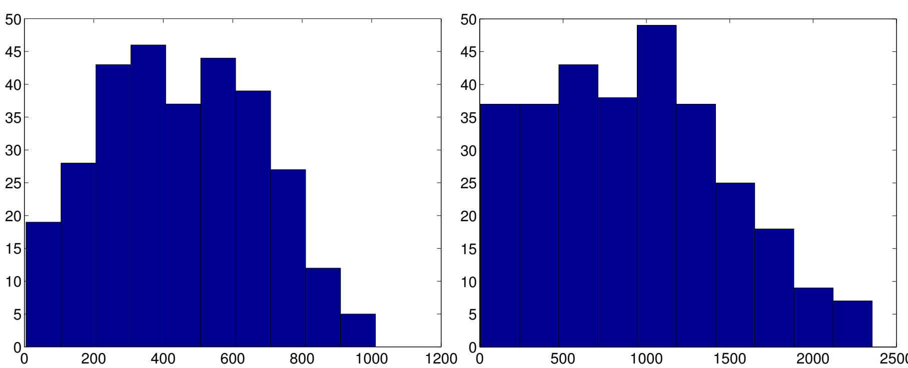
3.2.3. Įvertinių empirinis palyginimas

Mes nagrinėjome dvi skirtingas vienodo dydžio $N = 300$ tyrimo populiacijas. Pirmoji populiacija yra reali, paimta iš Lietuvos įmonių tyrimo ir turinti asimetrinį skirstinį (žr. 3.3 pav.). Antroji populiacija yra dirbtinė ir turinti skirstinį, artimą normaliajam (žr. 3.4 pav.).

Abi populiacijos buvo suskaidytos į du sluoksnius, atsižvelgiant į tyrimo kintamojo y didumą: populiacijos elementai, atitinkantys y reikšmes, mažesnes už kintamojo y vidurkį μ_y, sudaro pirmąjį sluoksnį, o likusieji elementai – antrąjį. Sluoksnių dydžius pažymėjome $N_h, h = 1, 2, N = N_1 + N_2$. Tyrime naudojome paprastąjį atsitiktinį sluoksninį ėmimą. Imties dydis $n = n_1 + n_2 = 150$ buvo paskirstytas į sluoksnius remiantis Neimano optimaliuoju paskirstymu. Taigi k-ojo populiacijos elemento priklausymo imčiai tikimybė yra $\pi_k = n_h/N_h$, jei k-asis elementas priklauso h-ajam sluoksniui. Atliekant eksperimentą buvo tariama, kad laisvai pasirenkami svoriai $q_k = 1$ visiems k.



3.3 pav. *Reali populiacija: tyrimo kintamųjų y (kairėje) ir z (dešinėje) reikšmių histogramos*



3.4 pav. *Dirbtinė populiacija: tyrimo kintamųjų y (kairėje) ir z (dešinėje) reikšmių histogramos*

3.3 lentelė. *Reali populiacija: populiacijos dispersijos vertinimas (tikroji populiacijos dispersijos reikšmė* $S_y^2 = 1\,196\,163$*, imties dydis* $n = 150$*)*

Įvertinys	Įvertis	Dispersija $\times 10^{-10}$	$POSL$	$VKP \times 10^{-10}$	cv	r_{min}	r_{max}
Koreliacijos koeficientas $\rho(y,a) = 0{,}9$							
$\widehat{S}_{yy1}^{(non)}$	850 215	5,582	$-345\,948$	17,550	0,2779	0,8731	2,2405
$\widehat{S}_{yy6}^{(non)}$	852 128	5,535	$-344\,035$	17,371	0,2761	0,8704	2,2562
$\widehat{S}_{yy1}^{(tot)}$	749 496	18,515	$-446\,667$	38,466	0,5741	0,8050	1,2230
$\widehat{S}_{yy6}^{(tot)}$	753 876	18,600	$-442\,287$	38,162	0,5721	0,8035	1,2357
$\widehat{S}_{yy1}^{(lin)}$	1 019 964	1,923	$-176\,199$	5,028	0,1360	0,8408	4,0086
$\widehat{S}_{yy6}^{(lin)}$	1 025 198	1,902	$-170\,965$	4,825	0,1345	0,8409	4,0742
$\widehat{S}_1^2$	909 482	42,565	$-286\,681$	50,784	0,7174	–	–
$\widehat{S}_2^2$	912 535	42,833	$-283\,628$	50,877	0,7172	–	–
Koreliacijos koeficientas $\rho(y,a) = 0{,}6$							
$\widehat{S}_{yy1}^{(non)}$	745 908	18,356	$-450\,255$	38,629	0,5744	0,8733	1,2369
$\widehat{S}_{yy6}^{(non)}$	747 976	18,430	$-448\,187$	38,518	0,5740	0,8737	1,2491
$\widehat{S}_{yy1}^{(tot)}$	813 280	34,003	$-382\,883$	48,663	0,7170	0,8629	1,1187
$\widehat{S}_{yy6}^{(tot)}$	816 014	34,037	$-380\,149$	48,488	0,7150	0,8608	1,1252
$\widehat{S}_{yy1}^{(lin)}$	760 013	13,145	$-436\,150$	32,167	0,4770	0,8698	1,5691
$\widehat{S}_{yy6}^{(lin)}$	762 915	13,179	$-433\,248$	31,949	0,4758	0,8710	1,5939
$\widehat{S}_1^2$	885 164	41,755	$-310\,999$	51,427	0,7300	–	–
$\widehat{S}_2^2$	888 139	42,018	$-308\,024$	51,506	0,7299	–	–
Koreliacijos koeficientas $\rho(y,a) = 0{,}3$							
$\widehat{S}_{yy1}^{(non)}$	751 833	29,408	$-444\,330$	49,151	0,7213	0,9179	1,0871
$\widehat{S}_{yy6}^{(non)}$	752 713	29,369	$-443\,450$	49,034	0,7200	0,9173	1,0913
$\widehat{S}_{yy1}^{(tot)}$	793 639	36,244	$-402\,524$	52,447	0,7586	0,8909	1,1092
$\widehat{S}_{yy6}^{(tot)}$	794 481	36,130	$-401\,682$	52,264	0,7566	0,8867	1,1149
$\widehat{S}_{yy1}^{(lin)}$	722 704	20,260	$-473\,459$	42,677	0,6228	0,9191	1,1737
$\widehat{S}_{yy6}^{(lin)}$	724 920	20,379	$-471\,243$	42,586	0,6227	0,9203	1,1836
$\widehat{S}_1^2$	812 395	36,580	$-383\,768$	51,307	0,7445	–	–
$\widehat{S}_2^2$	815 144	36,810	$-381\,019$	51,328	0,7443	–	–

Kiekvienam įvertiniui buvo išrinkta $m = 500$ imčių $\mathbf{s}_j$ ir galutiniu įverčiu laikomas aritmetinis gautųjų įverčių vidurkis. Empirinė įvertinių dispersija buvo taip pat apskaičiuota. Ji kokiam nors įvertiniui $\widehat{Cov}(y,z)$ apibrėžiama taip:

$$Var_{emp} = \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{Cov}_{\mathbf{s}_i} - \frac{1}{m}\sum_{j=1}^{m}\widehat{Cov}_{\mathbf{s}_j}\right)^2,$$

čia $\widehat{Cov}_{\mathbf{s}_k}$ yra kovariacijos įvertis, gautas naudojant $\mathbf{s}_k$ imties duomenis.

3.4 lentelė. *Dirbtinė populiacija: populiacijos dispersijos vertinimas (tikroji populiacijos dispersijos reikšmė $S_y^2 = 350\,262$, imties dydis $n = 150$)*

Įvertinys	Įvertis	Dispersija $\times 10^{-8}$	$POSL$	$VKP \times 10^{-8}$	cv	r_{min}	r_{max}
Koreliacijos koeficientas $\rho(y,a) = 0{,}8$							
$\widehat{S}_{yy1}^{(non)}$	346 587	9,031	$-3\,675$	9,166	0,086 7	0,893 8	1,115 0
$\widehat{S}_{yy6}^{(non)}$	346 812	8,913	$-3\,450$	9,032	0,086 1	0,897 3	1,106 6
$\widehat{S}_{yy1}^{(tot)}$	353 933	14,157	3 671	14,292	0,106 3	0,996 4	1,013 3
$\widehat{S}_{yy6}^{(tot)}$	353 664	14,136	3 402	14,252	0,106 3	0,995 3	1,011 0
$\widehat{S}_{yy1}^{(lin)}$	354 746	7,115	4 484	7,316	0,075 2	1,000 0	1,398 6
$\widehat{S}_{yy6}^{(lin)}$	355 090	6,926	4 828	7,159	0,074 1	1,000 0	1,335 9
$\widehat{S}_1^2$	350 137	15,762	-125	15,762	0,113 4	–	–
$\widehat{S}_2^2$	350 568	15,802	306	15,803	0,113 4	–	–
Koreliacijos koeficientas $\rho(y,a) = 0{,}6$							
$\widehat{S}_{yy1}^{(non)}$	351 361	16,130	1 099	16,142	0,114 3	0,868 7	1,109 4
$\widehat{S}_{yy6}^{(non)}$	351 945	15,750	1 683	15,778	0,112 8	0,883 1	1,096 4
$\widehat{S}_{yy1}^{(tot)}$	352 427	15,302	2 165	15,348	0,111 0	0,996 5	1,010 0
$\widehat{S}_{yy6}^{(tot)}$	352 214	15,275	1 952	15,314	0,111 0	0,995 6	1,008 4
$\widehat{S}_{yy1}^{(lin)}$	369 884	19,334	19 622	23,184	0,118 9	1,000 0	1,330 6
$\widehat{S}_{yy6}^{(lin)}$	369 410	18,780	19 148	22,447	0,117 3	1,000 0	1,276 7
$\widehat{S}_1^2$	353 414	17,347	3 152	17,446	0,117 8	–	–
$\widehat{S}_2^2$	353 853	17,391	3 591	17,520	0,117 9	–	–
Koreliacijos koeficientas $\rho(y,a) = 0{,}2$							
$\widehat{S}_{yy1}^{(non)}$	352 793	20,358	2 531	20,422	0,127 9	0,909 1	1,162 6
$\widehat{S}_{yy6}^{(non)}$	353 133	19,709	2 871	19,791	0,125 7	0,912 6	1,146 9
$\widehat{S}_{yy1}^{(tot)}$	349 222	17,321	$-1\,040$	17,332	0,119 2	0,997 3	1,006 5
$\widehat{S}_{yy6}^{(tot)}$	349 145	17,313	$-1\,117$	17,326	0,119 2	0,996 6	1,007 1
$\widehat{S}_{yy1}^{(lin)}$	384 713	20,420	34 451	32,288	0,117 5	1,000 0	1,398 4
$\widehat{S}_{yy6}^{(lin)}$	385 119	20,284	34 857	32,434	0,116 9	1,000 0	1,362 3
$\widehat{S}_1^2$	348 696	16,832	$-1\,566$	16,856	0,117 7	–	–
$\widehat{S}_2^2$	349 126	16,873	$-1\,136$	16,886	0,117 7	–	–

Pirmiausia nagrinėsime atskirą atvejį, kai $y = z$, t. y. vertinsime populiacijos dispersiją.

Kalibruotųjų svorių sklaida yra svarbi kalibravimo procedūros ypatybė. Tegul $w_k(j)$ žymi kalibruotus svorius, gautus iš s_j imties. Mes apskaičiuosime dvi svorių w_k sklaidos charakteristikas

3.5 lentelė. *Reali populiacija: populiacijos kovariacijos vertinimas (tikroji populiacijos kovariacijos reikšmė* $Cov(y,z) = 66\,083\,066$*, imties dydis* $n = 150$*)*

Įvertinys	Įvertis	Dispersija $\times 10^{-13}$	$POSL$	$VKP \times 10^{-13}$	cv	r_{min}	r_{max}
$\rho(y,a)=0{,}81$	$\rho(z,b)=0{,}90$	$\rho(y,b)=0{,}63$	$\rho(z,a)=0{,}60$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	62 592 219	3,307	−3 490 848	4,526	0,091 9	0,847 2	1,395 5
$\widehat{Cov}_{w6}^{(non)}(y,z)$	62 593 288	3,307	−3 489 778	4,525	0,091 9	0,846 1	1,392 8
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	59 978 869	5,754	−6 104 197	9,480	0,126 5	0,731 7	1,300 9
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	59 972 059	5,770	−6 111 007	9,505	0,126 7	0,734 4	1,301 7
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	66 629 202	2,900	546 136	2,930	0,080 8	0,745 3	1,823 8
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	66 618 323	2,893	535 256	2,922	0,080 7	0,740 1	1,818 8
$\widehat{Cov}_1(y,z)$	60 480 672	12,324	−5 602 394	15,463	0,183 6	–	–
$\widehat{Cov}_2(y,z)$	60 627 514	12,381	−5 455 553	15,357	0,183 5	–	–
$\rho(y,a)=0{,}21$	$\rho(z,b)=0{,}90$	$\rho(y,b)=0{,}63$	$\rho(z,a)=0{,}15$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	61 874 002	8,533	−4 209 064	10,305	0,149 3	0,770 7	1,301 6
$\widehat{Cov}_{w6}^{(non)}(y,z)$	61 866 490	8,544	−4 216 576	10,322	0,149 4	0,766 5	1,301 8
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	61 013 304	5,796	−5 069 762	8,366	0,124 8	0,747 0	1,238 9
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	60 997 722	5,816	−5 085 344	8,403	0,125 0	0,750 7	1,239 7
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	60 208 806	11,510	−5 874 261	14,961	0,178 2	0,725 2	1,215 3
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	60 201 992	11,509	−5 881 074	14,968	0,178 2	0,723 4	1,215 1
$\widehat{Cov}_1(y,z)$	61 000 722	12,326	−5 082 344	14,909	0,182 0	–	–
$\widehat{Cov}_2(y,z)$	61 148 158	12,380	−4 934 908	14,815	0,182 0	–	–
$\rho(y,a)=0{,}23$	$\rho(z,b)=0{,}31$	$\rho(y,b)=0{,}19$	$\rho(z,a)=0{,}16$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	61 392 219	15,133	−4 690 847	17,333	0,200 4	0,597 8	1,531 2
$\widehat{Cov}_{w6}^{(non)}(y,z)$	61 228 861	14,670	−4 854 206	17,026	0,197 8	0,577 7	1,539 4
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	60 520 481	12,505	−5 562 585	15,600	0,184 8	0,842 4	1,140 3
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	60 521 271	12,488	−5 561 796	15,582	0,184 6	0,844 9	1,139 8
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	60 170 358	12,457	−5 912 708	15,953	0,185 5	0,719 4	1,179 2
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	60 171 222	12,458	−5 911 844	15,953	0,185 5	0,723 4	1,178 4
$\widehat{Cov}_1(y,z)$	60 420 502	12,462	−5 662 564	15,668	0,184 8	–	–
$\widehat{Cov}_2(y,z)$	60 567 298	12,518	−5 515 768	15,561	0,184 7	–	–

$$r_{max} = \frac{1}{m}\sum_{j=1}^{m}\max_{k\in\mathbf{s}_j}\frac{w_k(j)}{d_k}, \quad r_{min} = \frac{1}{m}\sum_{j=1}^{m}\min_{k\in\mathbf{s}_j}\frac{w_k(j)}{d_k}.$$

Įvertinsime populiacijos dispersiją naudodamiesi dviem standartiniais popu-

3.6 lentelė. *Dirbtinė populiacija: populiacijos kovariacijos vertinimas (tikroji populiacijos kovariacijos reikšmė $Cov(y,z) = 67\,644$, imties dydis $n = 150$)*

Įvertinys	Įvertis	Dispersija $\times 10^{-6}$	$POSL$	$VKP \times 10^{-6}$	cv	r_{min}	r_{max}
$\rho(y,a) = 0{,}8$ $\rho(z,b) = 0{,}8$							
$\widehat{Cov}_{w1}^{(non)}(y,z)$	67 283	40,674	-361	40,805	0,094 8	0,971 1	1,028 0
$\widehat{Cov}_{w6}^{(non)}(y,z)$	67 280	40,682	-364	40,815	0,094 8	0,971 0	1,027 9
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	67 366	43,573	-278	43,650	0,098 0	0,873 8	1,122 0
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	67 367	43,545	-277	43,622	0,098 0	0,873 5	1,121 9
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	67 522	32,324	-122	32,339	0,084 2	0,948 4	1,053 5
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	67 521	32,315	-123	32,330	0,084 2	0,948 9	1,053 1
$\widehat{Cov}_1(y,z)$	67 282	43,932	-363	44,064	0,098 5	–	–
$\widehat{Cov}_2(y,z)$	67 339	44,039	-305	44,132	0,098 5	–	–

liacijos dispersijos įvertiniais:

$$\widehat{S}_1^2 = \widehat{Cov}_1(y,y) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} d_k(y_k - \widehat{\mu}_y)^2, \quad \widehat{\mu}_y = \frac{1}{N}\sum_{k\in\mathbf{s}} d_k y_k,$$

ir $\widehat{S}_2^2 = \widehat{Cov}_2(y,y)$. Atsižvelgdami į tai, kokia kalibravimo lygtis – (3.10), (3.11) ar (3.12) – naudojama, kalibruotuosius dispersijos įvertinius pažymėjome $\widehat{S}_{yyi}^{(non)}$, $\widehat{S}_{yyi}^{(lin)}$ ir $\widehat{S}_{yyi}^{(tot)}$. Čia indeksas i nurodo atstumo funkcijos L_i, kuri naudojama įvertiniui apibrėžti, numerį. Analogiškai pažymėti ir atitinkami populiacijos kovariacijos įvertiniai: $\widehat{Cov}_{wi}^{(non)}(y,z)$, $\widehat{Cov}_{wi}^{(lin)}(y,z)$, $\widehat{Cov}_{wi}^{(tot)}(y,z)$.

Lentelėse 3.3, 3.4, 3.5 ir 3.6 pateikiami matematinio modeliavimo rezultatai, priklausantys nuo keleto skirtingų papildomų kintamųjų, skirtingai koreliuojančių su tyrimo kintamaisiais. Čia taip pat yra įvertinti įverčių variacijos koeficientai.

Kalibruotų įvertinių, sukonstruotų naudojantis ta pačia kalibravimo lygtimi, tikslumas yra labai panašus, nepaisant naudojamos atstumo funkcijos. Dėl šios priežasties mes pateikėme rezultatus tik atstumo funkcijoms L_1 ir L_6, apibrėžtoms atitinkamai (1.12) bei (1.18) lygybėmis.

Jei tyrimo ir papildomų kintamųjų koreliacija yra stipri, tai kalibruoti įvertiniai yra efektyvesni, palyginti su tradiciniais įvertiniais, kurie nenaudoja papildomų kintamųjų. Tai dar labiau išryškėja atliekant eksperimentą su asimetrine populiacija. Kai koreliacija silpna, visi įvertiniai yra panašios kokybės. Kai papildomi ir tyrimo kintamieji yra gerai koreliuoti, tai tiesinis kalibravimas yra efektyviausias.

Tuo atveju, kai vertiname populiacijos kovariaciją ir turime vieną papildomą kintamąjį, gerai koreliuotą su atitinkamu tyrimo kintamuoju, o kitą papildomą kintamąjį, kuris blogai koreliuoja su juo susietu tyrimo kintamuoju, netiesinis kalibravimas tinka geriau nei tiesinis.

Įvertinių, sukonstruotų naudojant sumų kalibravimą, vidutinė kvadratinė paklaida VKP yra dvigubai didesnė nei įvertinių, gautų netiesinio kalibravimo metu, ir daugiau nei trigubai didesnė už vidutinę kvadratinę paklaidą VKP tų įvertinių, kurie gauti naudojant tiesinį kalibravimą.

3.2.4. Kalibruotieji kovariacijos įvertiniai, naudojantys keletą svorių sistemų

Nagrinėkime kitus, bendresnius baigtinės populiacijos kovariacijos įvertinius, sukonstruotus naudojant keletą svorių sistemų. Nauji kalibruotieji kovariacijos įvertiniai yra tokio pavidalo:

$$\widehat{Cov}_{mw}(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k^{[1]}\Big(y_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[2]} y_l\Big)\Big(z_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[3]} z_l\Big). \tag{3.16}$$

Čia kalibruotųjų svorių viršutiniai indeksai [1], [2], [3] žymi skirtingas svorių sistemas.

Nagrinėjant šiuos įvertinius, kalibruotiems svoriams apibrėžti taip pat gali būti naudojama keletas kalibravimo lygčių. Panagrinėkime kai kurias iš jų.

1 atvejis. Galima naudotis netiesine lygtimi

$$\widehat{Cov}_{mw}(a,b) = Cov(a,b). \tag{3.17}$$

2 atvejis. Svorių sistemas $w_k^{[1]}$, $w_k^{[2]}$, $w_k^{[3]}$ galima apibrėžti šiomis kalibravimo lygtimis:

$$\frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k^{[1]}\big(a_k-\mu_a\big)\big(b_k-\mu_b\big) = Cov(a,b), \tag{3.18}$$

$$\sum_{k\in\mathbf{s}} w_k^{[2]} a_k = t_a, \quad \sum_{k\in\mathbf{s}} w_k^{[3]} b_k = t_b. \tag{3.19}$$

3 atvejis. Pirmąją svorių sistemą $w_k^{[1]}$ galima apibrėžti naudojantis netiesinio kalibravimo (3.10) lygtimi, o atskirai kalibruojant sumas t_a ir t_b rasti antrąją ir trečiąją svorių sistemas $w_k^{[2]}$, $w_k^{[3]}$.

Šiais trimis atvejais galutinai kalibruotiems svoriams apibrėžti gali būti naudojama atstumo funkcija

$$L(w,d) = \alpha_1 \sum_{k\in\mathbf{s}} \frac{(w_k^{[1]} - d_k)^2}{d_k\, q_k} + \alpha_2 \sum_{k\in\mathbf{s}} \frac{(w_k^{[2]} - d_k)^2}{d_k\, q_k} + \alpha_3 \sum_{k\in\mathbf{s}} \frac{(w_k^{[3]} - d_k)^2}{d_k\, q_k}, \tag{3.20}$$

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

4 atvejis. Galima nagrinėti kovariacijos įvertinį, naudojantį dvi skirtingas svorių sistemas:

$$\widehat{Cov}_{mw}(y,z) = \frac{1}{N-1} \sum_{k\in\mathbf{s}} w_k^{[1]} \Big(y_k - \frac{1}{N} \sum_{l\in\mathbf{s}} w_l^{[2]} y_l \Big) \Big(z_k - \frac{1}{N} \sum_{l\in\mathbf{s}} w_l^{[2]} z_l \Big). \tag{3.21}$$

Pirmąją svorių sistemą $w_k^{[1]}$ apibrėžtų (3.18) lygtis, o antrąją gautume kalibruodami žinomas sumas t_a ir t_b:

$$\sum_{k\in\mathbf{s}} w_k^{[2]} a_k = t_a, \qquad \sum_{k\in\mathbf{s}} w_k^{[2]} b_k = t_b. \tag{3.22}$$

5 atvejis. Galima naudotis kita dviejų svorių sistemų kombinacija. Pirmąją svorių sistemą $w_k^{[1]}$ rastume naudodamiesi netiesinio kalibravimo (3.10) lygtimi, o antrąją – kalibruojant sumas t_a ir t_b bei naudojantis (3.22) lygtimis.

6 atvejis. Svorių sistema $w_k^{[1]}$ tenkina (3.18) lygtį, tuo tarpu svorių sistemos $w_k^{[2]}$ ir $w_k^{[3]}$ sutampa ir yra gaunamos naudojantis netiesinio kalibravimo (3.10) lygtimi.

Atstumo funkcija paskutiniais trimis atvejais gali būti tokia:

$$L'(w,d) = \beta \sum_{k\in\mathbf{s}} \frac{(w_k^{[1]} - d_k)^2}{d_k\, q_k} + (1-\beta) \sum_{k\in\mathbf{s}} \frac{(w_k^{[2]} - d_k)^2}{d_k\, q_k}, \quad 0 < \beta < 1. \tag{3.23}$$

Pirmas atvejis yra labiausiai komplikuotas analitine prasme. Naudojantis (3.17) lygtimi gaunamos labai sudėtingos apytikslių iteracinių kalibravimo lygties sprendinių išraiškos.

Kitame teiginyje kalbama apie (3.16) įvertinio svorius $w_k^{[1]}$, $w_k^{[2]}$, $w_k^{[3]}$ visais šešiais šiame skyrelyje minėtais kovariacijos įvertinių konstravimo atvejais. Minimizuodami (3.20) ir (3.23) atstumo funkcijas, imame šiuos koeficientus: $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$, $\beta = \frac{1}{2}$.

3.5 teiginys. *Svoriai* $w_k = w_k^{[i]}$, $k \in \mathbf{s}$, $i = 1, 2, 3$, *kurie tenkina* (3.17) *kalibravimo lygtį ir minimizuoja* (3.20) *atstumo funkciją, tenkina lygtį* $w_k^{[i]} = d_k u_k^{[i]}$. *Čia* $u_k^{[i]} = 1 + \lambda q_k e_k^{[i]}$,

$$
\begin{aligned}
e_k^{[1]} &= a_k b_k - a_k \widehat{\mu}_{bw^{[3]}} - b_k \widehat{\mu}_{aw^{[2]}} + \widehat{\mu}_{aw^{[2]}} \widehat{\mu}_{bw^{[3]}}, \\
e_k^{[2]} &= -a_k \Big(\widehat{\mu}_{bw^{[1]}} - \frac{\hat{N}_{w^{[1]}}}{N} \widehat{\mu}_{bw^{[3]}} \Big), \\
e_k^{[3]} &= -b_k \Big(\widehat{\mu}_{aw^{[1]}} - \frac{\hat{N}_{w^{[1]}}}{N} \widehat{\mu}_{aw^{[2]}} \Big), \\
\lambda &= \widehat{A} \Big(\sum_{k \in \mathbf{s}} d_k q_k (a_k b_k e_k^{[1]} + e_k^{[2]} - e_k^{[3]}) \Big)^{-1},
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A} = (N-1)Cov(a,b) + N \widehat{\mu}_{bw^{[3]}} \Big(\widehat{\mu}_{aw^{[1]}} - \frac{\hat{N}_{w^{[1]}}}{N} \widehat{\mu}_{aw^{[2]}} \Big) + N \widehat{\mu}_{aw^{[2]}} \widehat{\mu}_{bw^{[1]}} + \\
+ \hat{N}_{w^{[2]}} - \hat{N}_{w^{[3]}} - \sum_{k \in \mathbf{s}} d_k a_k b_k,
\end{aligned}
$$

$$
\begin{aligned}
&\widehat{\mu}_{aw^{[1]}} = \frac{1}{N} \sum_{k \in \mathbf{s}} w_k^{[1]} a_k, \qquad \widehat{\mu}_{aw^{[2]}} = \frac{1}{N} \sum_{k \in \mathbf{s}} w_k^{[2]} a_k \\
&\widehat{\mu}_{bw^{[1]}} = \frac{1}{N} \sum_{k \in \mathbf{s}} w_k^{[1]} b_k, \qquad \widehat{\mu}_{bw^{[3]}} = \frac{1}{N} \sum_{k \in \mathbf{s}} w_k^{[3]} b_k \\
&\hat{N}_{w^{[1]}} = \sum_{k \in \mathbf{s}} w_k^{[1]}, \qquad \hat{N}_{w^{[2]}} = \sum_{k \in \mathbf{s}} w_k^{[2]}, \qquad \hat{N}_{w^{[3]}} = \sum_{k \in \mathbf{s}} w_k^{[3]};
\end{aligned}
$$

2-*uoju,* 4-*uoju ir* 6-*uoju įvertinių konstravimo atvejais pirmoji svorių sistema* $w_k^{[1]}$ *apibrėžiama šiomis lygtimis:*

$$
w_k^{[1]} = d_k \Big(1 + q_k \Big(\sum_{l=1}^{N} x_l^{[1]} - \sum_{l \in \mathbf{s}} d_l x_l^{[1]} \Big) \Big(\sum_{l \in \mathbf{s}} d_l q_l (x_l^{[1]})^2 \Big)^{-1} x_k^{[1]} \Big),
$$

čia $x_k^{[1]} = (a_k - \mu_a)(b_k - \mu_b)$;

3-*uoju ir* 5-*uoju atvejais pirmoji svorių sistema, o* 6-*uoju atveju antroji ir trečioji svorių sistemos* $w_k^{[2]}$, $w_k^{[3]}$ *yra tiesiogiai apskaičiuojamos remiantis 3.3 teiginiu;*

2-uoju ir 3-uoju atvejais svoriai $w_k^{[2]}$ ir $w_k^{[3]}$ yra randami iš lygčių

$$w_k^{[i]} = d_k\Big(1 + q_k\Big(\sum_{l=1}^{N} x_l^{[i]} - \sum_{l\in\mathbf{s}} d_l x_l^{[i]}\Big)\Big(\sum_{l\in\mathbf{s}} d_l q_l (x_l^{[i]})^2\Big)^{-1} x_k^{[i]}\Big),$$

čia $x_k^{[2]} = a_k$, $x_k^{[3]} = b_k$, $i = 2, 3$;

4-uoju ir 5-uoju atvejais antroji svorių sistema $w_k^{[2]}$ tenkina šias lygtis:

$$w_k^{[2]} = d_k\Big(1 + q_k\Big(\sum_{l=1}^{N} \mathbf{x}_l^{[2]\prime} - \sum_{l\in\mathbf{s}} d_l \mathbf{x}_l^{[2]\prime}\Big)\Big(\sum_{l\in\mathbf{s}} d_l q_l \mathbf{x}_l^{[2]} \mathbf{x}_l^{[2]\prime}\Big)^{-1} \mathbf{x}_k^{[2]}\Big),$$

čia $\mathbf{x}_k^{[2]} = (a_k, b_k)'$.

Šio teiginio įrodymas yra analogiškas 3.3 teiginio įrodymui.

3.2.5. Matematinis modeliavimas: skirtingų svorių sistemų įtaka vertinimo tikslumui

Toliau atliksime matematinį modeliavimą, kurio metu palyginsime kalibruotus ir vieną svorių sistemą naudojančius kovariacijos įvertinius su įvertiniais, naudojančiais dvi ar tris svorių sistemas. Kalibruoti ir vieną svorių sistemą naudojantys įvertiniai, gauti taikant tą pačią kalibravimo lygtį, yra labai panašūs nepaisant atstumo funkcijos, todėl tyrime apsiribosime imdami tik tris šiuos kovariacijos įvertinius: $\widehat{Cov}_{w1}^{(non)}(y,z)$, $\widehat{Cov}_{w1}^{(tot)}(y,z)$, $\widehat{Cov}_{w1}^{(lin)}(y,z)$.

Kovariacijos įvertinius, atitinkančius 6 minėtus įvertinių su keliomis svorių sistemomis konstravimo atvejus, pažymėjome atitinkamai $\widehat{Cov}_{mw}^{(i)}(y,z)$, $i = 1, 2,$ 3, 4, 5, 6. Pavyzdžiui, $\widehat{Cov}_{mw}^{(1)}(y,z)$ žymimas įvertinys, naudojantis tokias tris svorių sistemas, kurios tenkina (3.17) kalibravimo lygtį ir minimizuoja (3.20) nuostolių funkciją.

Eksperimentas atliktas naudojant tą pačią populiaciją iš Lietuvos įmonių tyrimo (žr. 3.2.3 skyrelį) ir esant toms pačioms tyrimo sąlygoms. Taigi sudarant imtis naudotasi tuo pačiu paprastuoju atsitiktiniu sluoksniniu ėmimu, imties dydis $n = 150$, kiekvienam įvertiniui išrinkta po 500 imčių ir kiekvienai iš jų apskaičiuoti kovariacijos įverčiai. Naudojantis gautais įverčiais, įvertintos pagrindinės įvertinių tikslumo charakteristikos (žr. 3.7 lentelę).

Iš lentelės pirmos dalies (kai $\rho(y,a) = 0{,}81$ ir $\rho(z,b) = 0{,}90$) matyti, kad taikant tiesinio ir netiesinio kalibravimo kombinaciją, kai pirmoji svorių sistema

3.7 lentelė. *Pagrindinės įvertintos kovariacijos įvertinių tikslumo charakteristikos (tikroji kovariacijos reikšmė $Cov(y, z) = 66\,083\,066$, imties dydis $n = 150$)*

Įvertinys	Įvertis	Dispersija $\times\, 10^{-13}$	Poslinkis	$VKP \times 10^{-13}$	cv
$\rho(y,a) = 0{,}81$ $\rho(z,b) = 0{,}90$ $\rho(y,b) = 0{,}63$ $\rho(z,a) = 0{,}60$					
$\widehat{Cov}_{w1}^{(non)}(y,z)$	62 808 825	2,749	-3 274 241	3,821	0,083 5
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	60 824 495	5,313	-5 258 571	8,079	0,119 8
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	65 654 598	2,212	-428 467	2,231	0,071 6
$\widehat{Cov}_{mw}^{(1)}(y,z)$	65 957 347	2,165	-125 718	2,166	0,070 5
$\widehat{Cov}_{mw}^{(2)}(y,z)$	65 761 298	2,119	-321 767	2,130	0,070 0
$\widehat{Cov}_{mw}^{(3)}(y,z)$	62 712 647	2,804	-3 370 419	3,940	0,084 4
$\widehat{Cov}_{mw}^{(4)}(y,z)$	65 777 778	2,121	-305 287	2,130	0,070 0
$\widehat{Cov}_{mw}^{(5)}(y,z)$	62 744 104	2,792	-3 338 961	3,907	0,084 2
$\widehat{Cov}_{mw}^{(6)}(y,z)$	65 753 625	2,107	-329 440	2,117	0,069 8
$\widehat{Cov}_1(y,z)$	61 224 682	10,386	-4 858 383	12,746	0,166 5
$\rho(y,a) = 0{,}21$ $\rho(z,b) = 0{,}90$ $\rho(y,b) = 0{,}63$ $\rho(z,a) = 0{,}15$					
$\widehat{Cov}_{w1}^{(non)}(y,z)$	61 888 467	6,741	-4 194 599	8,500	0,132 7
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	61 173 622	5,211	-4 909 444	7,622	0,118 0
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	60 411 548	9,494	-5 671 518	12,710	0,161 3
$\widehat{Cov}_{mw}^{(1)}(y,z)$	60 847 498	9,825	-5 235 568	12,566	0,162 9
$\widehat{Cov}_{mw}^{(2)}(y,z)$	60 702 528	9,378	-5 380 538	12,273	0,159 5
$\widehat{Cov}_{mw}^{(3)}(y,z)$	61 834 269	6,742	-4 248 797	8,547	0,132 8
$\widehat{Cov}_{mw}^{(4)}(y,z)$	60 900 339	9,204	-5 182 727	11,890	0,157 5
$\widehat{Cov}_{mw}^{(5)}(y,z)$	61 991 693	6,647	-4 091 373	8,321	0,131 5
$\widehat{Cov}_{mw}^{(6)}(y,z)$	60 762 008	9,444	-5 321 058	12,275	0,159 9
$\widehat{Cov}_1(y,z)$	61 207 562	9,776	-4 875 504	12,153	0,161 5
$\rho(y,a) = 0{,}23$ $\rho(z,b) = 0{,}31$ $\rho(y,b) = 0{,}19$ $\rho(z,a) = 0{,}16$					
$\widehat{Cov}_{w1}^{(non)}(y,z)$	61 937 538	12,133	-4 145 529	13,852	0,177 8
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	61 440 563	10,291	-4 642 503	12,446	0,165 1
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	61 014 951	10,291	-5 068 115	12,859	0,166 3
$\widehat{Cov}_{mw}^{(1)}(y,z)$	61 034 015	10,292	-5 049 051	12,841	0,166 2
$\widehat{Cov}_{mw}^{(2)}(y,z)$	61 038 828	10,282	-5 044 238	12,827	0,166 1
$\widehat{Cov}_{mw}^{(3)}(y,z)$	61 679 098	11,425	-4 403 968	13,365	0,173 3
$\widehat{Cov}_{mw}^{(4)}(y,z)$	61 080 625	10,300	-5 002 441	12,802	0,166 2
$\widehat{Cov}_{mw}^{(5)}(y,z)$	61 720 328	11,442	-4 362 738	13,345	0,173 3
$\widehat{Cov}_{mw}^{(6)}(y,z)$	61 312 037	10,369	-4 771 029	12,645	0,166 1
$\widehat{Cov}_1(y,z)$	61 257 357	10,260	-4 825 709	12,589	0,165 4

$w_k^{[1]}$ apibrėžiama (3.18) lygtimi, o antroji ir trečioji sistemos $w_k^{[2]}$, $w_k^{[3]}$ – (3.10) lygtimi, gauti geri rezultatai – įvertinys $\widehat{Cov}_{mw}^{(6)}$ yra tiksliausias. Jei pirmą svorių sistemą apibrėžtume netiesine kalibravimo lygtimi, o likusias dvi svorių sistemas – naudodami žinomų sumų kalibravimą, tai gautume įvertinius $\widehat{Cov}_{mw}^{(3)}$ ir $\widehat{Cov}_{mw}^{(5)}$, kurie dėl didelio poslinkio turi didesnę VKP nei kai kurie kalibruoti ir vieną svorių sistemą naudojantys kovariacijos įvertiniai. Įvertinių $\widehat{Cov}_{mw}^{(1)}$, $\widehat{Cov}_{mw}^{(2)}$, $\widehat{Cov}_{mw}^{(4)}$ tikslumas nežymiai skiriasi nuo įvertinio $\widehat{Cov}_{mw}^{(6)}$. Mažiausią poslinkį turi pirmas įvertinys $\widehat{Cov}_{mw}^{(1)}$.

Turint tik vieną gerai koreliuotą papildomą kintamąjį, t. y. kai $\rho(y,a) = 0{,}21$ ir $\rho(z,b) = 0{,}90$, įvertiniai $\widehat{Cov}_{mw}^{(3)}$ ir $\widehat{Cov}_{mw}^{(5)}$ yra tiksliausi nagrinėtų kelias svorių sistemas naudojančių įvertinių grupėje. Jų ir netiesinio įvertinio $\widehat{Cov}_{w1}^{(non)}$ atitinkamos tikslumo charakteristikos yra labai panašios. Tai galima būtų paaiškinti tuo, kad įvertinių $\widehat{Cov}_{mw}^{(3)}$ $\widehat{Cov}_{mw}^{(5)}$ pirmoji svorių sistema $w_k^{[1]}$ apibrėžta naudojantis netiesinio kalibravimo (3.10) lygtimi. Kitų kelias svorių sistemas naudojančių įvertinių ir įvertinio $\widehat{Cov}_{w1}^{(lin)}$, naudojančio vieną svorių sistemą, apibrėžiamą tiesinio kalibravimo lygtimi, vertinimo efektyvumas yra labai panašus. Šį faktą 2-uoju, 4-uoju ir 6-uoju kelias svorių sistemas naudojančių įvertinių atvejais galima būtų paaiškinti tiesinio kalibravimo (3.18) lygtimi, kuria buvo naudotasi pirmajai kalibruotų svorių sistemai $w_k^{[1]}$ apibrėžti.

Tuo atveju, kai tyrimo ir atitinkamų papildomų kintamųjų koreliacija yra silpna, visų kalibruotų įvertinių ir standartinio įvertinio $\widehat{Cov}_1$ tikslumas yra panašus. Imant blogai su tyrimo kintamaisiais koreliuotus papildomus kintamuosius apčiuopiamų rezultatų negauta. Standartinis įvertinys yra kur kas paprastesnis, todėl mes jį ir rekomenduotume naudoti vertinant populiacijos kovariaciją ir neturint gerai koreliuotų papildomų kintamųjų.

3.3. Kalibruotųjų populiacijos kovariacijos įvertinių dispersijos vertinimas

Šiame poskyryje nagrinėsime, kaip vertinama pateiktų baigtinės populiacijos kovariacijos kalibruotų įvertinių dispersija. Kaip jau buvo parodyta, jei koreliacija tarp tyrimo ir papildomų kintamųjų yra didelė, tai populiacijos kovariacijos kalibruotieji įvertiniai yra efektyvesni nei pagrįstieji imties planu. Esant silpnai tyrimo ir papildomų kintamųjų koreliacijai, visi įvertiniai yra panašios kokybės. Jau matėme, kad yra daug būdų kalibruotiems kovariacijos įvertiniams sukonstruoti. Aukščiau nagrinėjome trijų tipų įvertinius, kurie naudoja vieną svorių sistemą. Šių įvertinių dispersijų skaičiavimas tampa komplikuotas vien dėl to, kad turime labai sudėtingas kalibruotų svorių išraiškas. Pavyzdžiui, netiesinio kalibravimo atveju kalibruotus svorius apibrėžia neišreikštinės lygtys, todėl neegzistuoja išreikštinė analizinė įvertinio dispersijos išraiška. Taigi tokiais atvejais įvertinių dispersijoms vertinti taikomas grubus Teiloro ištiesinimo arba pseudo ištiesinimo metodas, kurį aprašysime kitame skyrelyje. Kai tik pateiksime kovariacijos įvertinių dispersijų įvertinius, atliksime matematinį modeliavimą ir palyginsime sukonstruotus įvertinių dispersijos įvertinius su empirine dispersija bei visrakčio dispersijos įvertiniu.

3.3.1. Kovariacijos įvertinių dispersijos vertinimas: grubus Teiloro ir pseudo ištiesinimo metodai

Yra keletas būdų, kaip įvertinti sudėtingų įvertinių dispersiją. Tai galėtų būti analiziniai (pvz., įvertinių ištiesinimo metodas) ar poimčių (pvz., visrakčio, savirankos[1]) metodai. Pasirinkus ištiesinimo techniką, užrašoma įvertinio tiesinė aproksimacija ir išvedama jos dispersijos išraiška, kuri naudojama įvertinio dispersijos įvertiniui konstruoti.

Mūsų atveju taikant analizinius dispersijos vertinimo metodus, susiduriama su tam tikrais sunkumais: netiesinio kalibravimo atveju nėra išreikštinių kalibruotų svorių w_k išraiškų ir patys įvertiniai yra daug sudėtingesni nei, pavyzdžiui, populiacijos sumos įvertiniai.

Sukonstruosime du grubius nagrinėtų kalibruotų kovariacijos įvertinių dispersijos įvertinius. Tuos pačius supaprastintus dispersijos įvertinius pritaikysime visiems kalibruotų įvertinių tipams.

Ištiesintasis įvertinys.

Pirmasis kalibruotųjų ir vieną svorių sistemą naudojančių kovariacijos įvertinių dispersijos įvertinys yra tokio pavidalo:

$$\widehat{Var}(\widehat{Cov}_w(y,z)) = \left(\frac{1}{N-1}\right)^2 \sum_{k,l\in\mathbf{s}} \left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\frac{\hat{e}_k}{\pi_k}\frac{\hat{e}_l}{\pi_l}, \tag{3.24}$$

čia

$$\hat{e}_k = \frac{w_k}{d_k}\left((y_k - \widehat{\mu}_{yw})(z_k - \widehat{\mu}_{zw}) + \left(\frac{\hat{N}_w}{N} - 1\right)(y_k\widehat{\mu}_{zw} + z_k\widehat{\mu}_{yw})\right).$$

[1] Angl. bootstrap method.

Šio įvertinio motyvacija yra tokia. Įvertinys $\widehat{Cov}_w(y,z)$, apibrėžtas (3.9) išraiška, gali būti užrašytas

$$\widehat{Cov}_w(y,z) = \frac{1}{N-1}\Big(\hat{t}_{yzw} - \frac{2}{N}\hat{t}_{yw}\hat{t}_{zw} + \frac{1}{N^2}\hat{t}_{yw}\hat{t}_{zw}\widehat{N}_w\Big),$$

čia

$$\begin{aligned} \hat{t}_{yzw} &= \sum_{k\in\mathbf{s}} w_k y_k z_k, \qquad & \hat{t}_{yw} &= \sum_{k\in\mathbf{s}} w_k y_k, \\ \hat{t}_{zw} &= \sum_{k\in\mathbf{s}} w_k z_k, \qquad & \widehat{N}_w &= \sum_{k\in\mathbf{s}} w_k. \end{aligned}$$

Pažymėkime

$$\begin{aligned} t_{yzw} &= \sum_{k=1}^{N} \frac{w_k}{d_k} y_k z_k, \qquad & t_{yw} &= \sum_{k=1}^{N} \frac{w_k}{d_k} y_k, \\ t_{zw} &= \sum_{k=1}^{N} \frac{w_k}{d_k} z_k, \qquad & N_w &= \sum_{k=1}^{N} \frac{w_k}{d_k}. \end{aligned}$$

Taško $(\hat{t}_{yzw}, \hat{t}_{yw}, \hat{t}_{zw}, \hat{N}_w) = (t_{yzw}, t_{yw}, t_{zw}, N_w)$ aplinkoje išskleiskime funkciją $\widehat{Cov}_w(y,z)$ Teiloro eilute ir skleidinio tiesinę dalį

$$\widehat{Cov}_{wl}(y,z) = \frac{1}{N-1}\Bigg(\hat{t}_{yzw} + \frac{t_{zw}}{N}\Big(\frac{N_w}{N} - 2\Big)\hat{t}_{yw} + \frac{t_{yw}}{N}\Big(\frac{N_w}{N} - 2\Big)\hat{t}_{zw} +$$

$$+\frac{t_{yw}t_{zw}}{N^2}\widehat{N}_w + \frac{2t_{yw}t_{zw}}{N}\Big(1 - \frac{N_w}{N}\Big)\Bigg).$$

pavadinkime *ištiesintuoju įvertiniu*.

Ištiesintojo įvertinio dispersija yra

$$Var(\widehat{Cov}_{wl}(y,z)) = \Big(\frac{1}{N-1}\Big)^2 Var\Big(\sum_{k\in\mathbf{s}} d_k e_k\Big),$$

čia

$$e_k = \frac{w_k}{d_k}\left(y_k z_k + \frac{t_{zw}}{N}\left(\frac{N_w}{N} - 2\right)y_k + \frac{t_{yw}}{N}\left(\frac{N_w}{N} - 2\right)z_k + \frac{1}{N^2}t_{yw}t_{zw}\right).$$

Darant prielaidą, kad kalibruoti svoriai nėra atsitiktiniai, ir remiantis 2.8.1 rezultatu iš [85], gaunama dispersijos išraiška

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

jos įvertiniu laikant šią statistiką:

$$\widehat{Var}(\widehat{Cov}_{wl}(y,z)) = \left(\frac{1}{N-1}\right)^2 \sum_{k,l\in\mathbf{s}} \left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\frac{\hat{e}_k}{\pi_k}\frac{\hat{e}_l}{\pi_l}. \qquad (3.25)$$

Reikšmės $\hat{e}_k$ apibrėžiamos pakeičiant nežinomus parametrus t_{yw}, t_{zw}, N_w jų įverčiais $\hat{t}_{yw}, \hat{t}_{zw}, \widehat{N}_w$.

Kalibruotųjų svorių sistemą, gautą naudojant netiesinį, tiesinį ar sumų kalibravimą, įrašius į (3.25) išraišką, gaunamas kovariacijos įvertinio, atitinkančio pasirinktą kalibravimo tipą ir atstumo funkciją, dispersijos įvertinys.

Apibrėžkime kitą dispersijos įvertinį. Kovariacijos (3.9) įvertinys gali būti išreikštas šiuo pavidalu:

$$\widehat{Cov}_w(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} d_k e_k^{(p)},$$

čia d_k yra imties plano svoriai,

$$e_k^{(p)} = \frac{w_k}{d_k}(y_k - \widehat{\mu}_{yw})(z_k - \widehat{\mu}_{zw}).$$

Naudodamiesi tuo pačiu 2.8.1 rezultatu iš [85], gauname antrąjį dispersijos įvertinį:

$$\widehat{Var}_p(\widehat{Cov}_w(y,z)) = \left(\frac{1}{N-1}\right)^2 \sum_{k,l\in\mathbf{s}} \left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\frac{e_k^{(p)}}{\pi_k}\frac{e_l^{(p)}}{\pi_l}. \qquad (3.26)$$

Jį pavadinkime *pseudo ištiesintuoju* įvertiniu.

Pirmojo dispersijos įvertinio (3.24) išraiška yra daug sudėtingesnė nei antrojo, tačiau sunku dabar pasakyti, kuriam iš šių įvertinių teikti pirmenybę.

Kaip jau minėjome, poimčių metodai, tokie kaip visrakčio, savirankos, taip pat gali būti naudojami sudėtingų įvertinių dispersijoms vertinti. Kitame skyrelyje empiriškai palyginsime abu kovariacijos įvertinių dispersijos įvertinius su visrakčio dispersijos įvertiniu ir su empirine dispersija.

3.3.2. Matematinio modeliavimo rezultatai

Eksperimente naudojome tą pačią realią ankstesnio pavyzdžio populiaciją (žr. 3.2.3 skyrelį). Populiacija suskaidyta į du sluoksnius atsižvelgiant į tyrimo kintamojo y didumą: populiacijos elementai, atitinkantys y reikšmes, mažesnes už kintamojo y vidurkį μ_y, sudaro pirmą sluoksnį, o likusieji elementai – antrąjį. Sluoksnių dydžius pažymėjome $N_h, h = 1, 2, N = N_1 + N_2$. Imtys renkamos naudojantis paprastuoju atsitiktiniu sluoksniniu ėmimu. Imties dydis $n = n_1 + n_2 = 100$ buvo paskirstytas į sluoksnius, remiantis Neimano optimaliuoju paskirstymu. Taigi populiacijos k-ojo elemento priklausymo imčiai tikimybė yra lygi $\pi_k = n_h/N_h$, jei k-asis elementas priklauso h-ajam sluoksniui. Eksperimentuojant buvo tariama, kad su visais k papildomi svoriai $q_k = 1$.

Išrinkome $m = 500$ nepriklausomų imčių $\mathbf{s}_j$ ir apskaičiavome populiacijos kovariacijos įvertinių empirinę dispersiją bei ištiesintojo (3.24) dispersijos įvertinio, pseudo ištiesintojo (3.26) dispersijos įvertinio, visrakčio dispersijos įvertinio įverčių aritmetinius vidurkius.

Norėdami parodyti papildomų kintamųjų privalumus, apskaičiavome standartinių populiacijos kovariacijos (3.8) įvertinių empirinę ir visrakčio dispersijas. Eksperimente naudojome tris skirtingas papildomų kintamųjų a ir b poras. Pirmu atveju abu papildomi kintamieji a ir b yra gerai koreliuoti su tyrimo kintamaisiais y ir z; antrame papildomų kintamųjų rinkinyje turime tik vieną gerai koreliuotą su atitinkamu tyrimo kintamuoju papildomą kintamąjį; trečiu atveju koreliacija tarp papildomų ir tyrimo kintamųjų yra silpna. Pateiktoje 3.8 lentelėje yra matematinio modeliavimo rezultatai, kurie priklauso nuo įvairių papildomų kintamųjų, skirtingai koreliuojančių su tyrimo kintamaisiais.

Kaip jau minėta, atsižvelgiant į tai, kokia – (3.10), (3.11) ar (3.12) – kalibravimo lygtimi naudojamasi, kalibruoti kovariacijos įvertiniai pažymėti taip:

$$\widehat{Cov}_{wi}^{(non)}(y,z), \quad \widehat{Cov}_{wi}^{(lin)}(y,z), \quad \widehat{Cov}_{wi}^{(tot)}(y,z).$$

Čia indeksas i rodo įvertiniui apibrėžti naudojamos atstumo funkcijos L_i numerį.

Kalibruoti įvertiniai, sukonstruoti naudojantis ta pačia kalibravimo lygtimi, yra labai panašaus tikslumo, nepaisant naudojamos atstumo funkcijos. Dėl šios priežasties mes atlikome skaičiavimus tik atstumo funkcijoms L_1 ir L_6.

3.8 lentelė. *Kalibruotų kovariacijos įvertinių dispersijos įverčiai ir empirinė dispersija (reali populiacija, $N = 300$, imties dydis: $n = 100$)*

Įvertinys	Empirinė dispersija $\times 10^{-13}$	Ištiesintas dispersijos įvertinys $\times 10^{-13}$	Pseudo ištiesintas dispersijos įvertinys $\times 10^{-13}$	Visrakčio dispersija $\times 10^{-13}$
$\rho(y,a) = 0{,}8$ $\rho(z,b) = 0{,}9$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	2,960	5,729	5,541	2,603
$\widehat{Cov}_{w6}^{(non)}(y,z)$	2,967	5,730	5,524	2,601
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	5,458	5,672	5,634	5,012
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	5,539	5,675	5,634	4,980
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	2,310	6,157	5,663	2,480
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	2,274	6,128	5,616	2,419
$\widehat{Cov}_1(y,z)$	9,840	—	—	7,878
$\widehat{Cov}_2(y,z)$	9,876	—	—	7,907
$\rho(y,a) = 0{,}2$ $\rho(z,b) = 0{,}9$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	7,052	7,485	7,478	5,682
$\widehat{Cov}_{w6}^{(non)}(y,z)$	7,107	7,481	7,471	5,742
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	4,871	5,654	5,616	5,042
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	5,002	5,675	5,637	5,032
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	10,029	7,267	7,315	7,869
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	10,018	7,254	7,309	7,866
$\widehat{Cov}_1(y,z)$	10,376	—	—	7,878
$\widehat{Cov}_2(y,z)$	10,411	—	—	7,907
$\rho(y,a) = 0{,}2$ $\rho(z,b) = 0{,}3$				
$\widehat{Cov}_{w1}^{(non)}(y,z)$	11,555	7,403	7,228	7,745
$\widehat{Cov}_{w6}^{(non)}(y,z)$	18,823	9,480	7,544	7,899
$\widehat{Cov}_{w1}^{(tot)}(y,z)$	10,023	6,975	6,943	7,610
$\widehat{Cov}_{w6}^{(tot)}(y,z)$	10,031	6,997	6,969	7,536
$\widehat{Cov}_{w1}^{(lin)}(y,z)$	10,409	7,206	7,249	7,950
$\widehat{Cov}_{w6}^{(lin)}(y,z)$	10,416	7,199	7,242	7,938
$\widehat{Cov}_1(y,z)$	10,306	—	—	7,878
$\widehat{Cov}_2(y,z)$	10,398	—	—	7,907

Matematinis modeliavimas rodo, kad jei tyrimo ir papildomų kintamųjų koreliacija yra didelė, tai įvertiniai, sukonstruoti naudojant tiesinį kalibravimą, turi mažiausią dispersiją. Tuo tarpu įvertinių, gautų naudojant sumų kalibravimą, efektyvumas yra mažiausias, nors tokių įvertinių nagrinėjimas yra motyvuotas imčių tyrimo praktika, kai tie patys kalibruoti svoriai naudojami visiems reikalingiems parametrams įvertinti.

Abiejų ištiesintųjų (3.24) ir (3.26) dispersijos įvertinių elgesys visais nagrinėtais atvejais yra labai panašus.

Visrakčio dispersijos įvertinys atrodo labiau adaptyvus – jis yra arčiau empirinės dispersijos ir geriau atspindi realią situaciją. Mes siūlome naudoti jį vertinant kalibruotųjų kovariacijos įvertinių dispersiją. Ištiesintasis (3.24) dispersijos įvertinys ir pseudo ištiesintas (3.26) įvertinys gali būti naudojami tik tuo atveju, kai tyrėjui pakanka grubaus ir apytikslio dispersijos vertinimo. Šie įvertiniai yra paprasti ir reikalauja daug mažiau skaičiavimo laiko, palyginti su visrakčio įvertiniu. Baigdami šį skyrelį, pirmenybę vis dėlto teikiame antrajam (3.26) dispersijos įvertiniui, nes jis yra paprastesnis ir gauti įverčiai yra arčiau empirinės ir visrakčio įvertinio dispersijų.

3.3.3. Tikslesnieji dispersijos įvertiniai

Ankstesniame skyrelyje buvo nagrinėti gana grubūs kovariacijos įvertinių dispersijos įvertiniai, gauti padarius prielaidą, kad kalibruotieji svoriai w_k nėra atsitiktiniai.

Kai kuriais atvejais, pavyzdžiui, naudojant tiesinio ar žinomų sumų kalibravimo lygtis bei atstumo funkciją L_1, svorius w_k galima gauti išreikštinio pavidalo. Todėl šiais atvejais įmanoma sukonstruoti tikslesnius kovariacijos įvertinių dispersijos įvertinius. Tai ir aptarsime šiame skyrelyje.

3.6 teiginys. *Kalibruotojo kovariacijos įvertinio*

$$\widehat{Cov}_{w1}^{(tot)}(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k\Big(y_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i y_i\Big)\Big(z_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i z_i\Big),$$

kurio svoriai w_k yra randami kalibruojant žinomas sumas t_a, t_b ir minimizuojant atstumo funkciją L_1, apibrėžtą (1.12) *lygybe, apytikslė dispersija yra*

$$AVar(\widehat{Cov}_{w1}^{(tot)}(y,z)) = \frac{1}{(N-1)^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}e_k e_l,$$

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

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

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

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

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

Įrodymas. *Baigtinės populiacijos kovariacijos įvertinį*

$$\widehat{Cov}_{w1}^{(tot)}(y, z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k\Big(y_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i y_i\Big)\Big(z_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i z_i\Big) \quad (3.27)$$

galima užrašyti

$$\widehat{Cov}_{w1}^{(tot)}(y, z) = \frac{1}{N-1}\bigg(\sum_{k\in\mathbf{s}} w_k y_k z_k + \frac{1}{N}\sum_{k\in\mathbf{s}} w_k y_k \sum_{k\in\mathbf{s}} w_k z_k\Big(\frac{1}{N}\sum_{k\in\mathbf{s}} w_k - 2\Big)\bigg).$$

Suma $\hat{N}_w = \sum_{k\in\mathbf{s}} w_k \approx N$. *Po šios aproksimacijos nagrinėjame naują kovariacijos įvertinį, nedaug tesiskiriantį nuo prieš tai turėto* (3.27)*:*

$$\widehat{Cov}_0(y, z) = \frac{1}{N-1}\Big(\sum_{k\in\mathbf{s}} w_k y_k z_k - \frac{1}{N}\sum_{k\in\mathbf{s}} w_k y_k \sum_{k\in\mathbf{s}} w_k z_k\Big). \quad (3.28)$$

Šio įvertinio dispersija apytiksliai yra lygi (3.27) *įvertinio dispersijai.*

Remiantis 3.1 teiginiu, svoriai w_k*, kurie tenkina kalibravimo* (3.12) *lygtis ir minimizuoja atstumo funkciją* L_1*, yra lygūs*

$$w_k = d_k\bigg(1 + \Big(\sum_{i=1}^{N}\mathbf{x}_i^{'} - \sum_{i\in\mathbf{s}} d_i\mathbf{x}_i^{'}\Big)\Big(\sum_{i\in\mathbf{s}} d_i q_i \mathbf{x}_i\mathbf{x}_i^{'}\Big)^{-1} q_k\mathbf{x}_k\bigg). \quad (3.29)$$

Populiacijos kovariacijos kalibruotus įvertinius konstravome turėdami tik du papildomus kintamuosius a *ir* b*, todėl* $\mathbf{x}_k$ *yra dvimatis vektorius* $\mathbf{x}_k = (a_k, b_k)'$*.*

Įrašę (3.29) *svorius į* (3.28) *išraišką, gauname*

$$\begin{aligned}\widehat{Cov}_0(y,z) = \frac{1}{N-1}\Big(&\sum_{k\in\mathbf{s}} d_k y_k z_k + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\big(\sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}\big)^{-1} \sum_{k\in\mathbf{s}} d_k q_k y_k z_k \mathbf{x}_k \\ &- \frac{1}{N}\big(\sum_{k\in\mathbf{s}} d_k y_k + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\big(\sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}\big)^{-1} \sum_{k\in\mathbf{s}} d_k q_k y_k \mathbf{x}_k\big) \\ &\times \big(\sum_{k\in\mathbf{s}} d_k z_k + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\big(\sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}\big)^{-1} \sum_{k\in\mathbf{s}} d_k q_k z_k \mathbf{x}_k\big)\Big),\end{aligned}$$

čia $\mathbf{t}_{\mathbf{x}} = \sum_{k=1}^{N} \mathbf{x}_k,\ \hat{\mathbf{t}}_{\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k \mathbf{x}_k.$
Pažymėkime

$$\begin{aligned}&\hat{t}_{yz} = \sum_{k\in\mathbf{s}} d_k y_k z_k, \quad \hat{t}_y = \sum_{k\in\mathbf{s}} d_k y_k, \quad \hat{t}_z = \sum_{k\in\mathbf{s}} d_k z_k, \\ &\hat{\mathbf{t}}_{yz\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k y_k z_k \mathbf{x}_k, \quad \hat{\mathbf{t}}_{y\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k y_k \mathbf{x}_k, \quad \hat{\mathbf{t}}_{z\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k z_k \mathbf{x}_k, \\ &\widehat{\mathbf{A}} = \sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}.\end{aligned}$$

Tuomet įvertinys $\widehat{Cov}_0(y,z)$ *įgyja pavidalą*

$$\begin{aligned}\widehat{Cov}_0(y,z) = \frac{1}{N-1}\Big(&\hat{t}_{yz} + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{yz\mathbf{x}} \\ &- \frac{1}{N}\big(\hat{t}_y + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}\big)\big(\hat{t}_z + \big(\mathbf{t}_{\mathbf{x}}^{'} - \hat{\mathbf{t}}_{\mathbf{x}}^{'}\big)\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}}\big)\Big) \\ = f(&\hat{t}_{yz}, \hat{t}_y, \hat{t}_z, \hat{\mathbf{t}}_{yz\mathbf{x}}, \hat{\mathbf{t}}_{y\mathbf{x}}, \hat{\mathbf{t}}_{z\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}}, \widehat{\mathbf{A}}).\end{aligned}$$

Taigi $\widehat{Cov}_0(y,z)$ *yra netiesinė funkcija, priklausanti nuo įvertinių* $\hat{t}_{yz}$, $\hat{t}_y$, $\hat{t}_z$, *vienuolikos įvertinių* $\hat{t}_{yzj}$, $\hat{t}_{yj}$, $\hat{t}_{zj}$, $\hat{t}_j$ *ir* $\hat{t}_{ij}$, *kurie yra atitinkamai dvimačių vektorių* $\hat{\mathbf{t}}_{yz\mathbf{x}}$, $\hat{\mathbf{t}}_{y\mathbf{x}}$, $\hat{\mathbf{t}}_{z\mathbf{x}}$, $\hat{\mathbf{t}}_{\mathbf{x}}$ *bei matricos* $\widehat{\mathbf{A}}$ *komponentės.*

Ištiesinsime funkciją $\widehat{Cov}_0(y,z)$ *taikydami Teiloro ištiesinimo metodą. Šiam tikslui reikalingos šios dalinės išvestinės:*

$$
\begin{aligned}
&\frac{\partial f}{\partial \hat{t}_{yz}} = \frac{1}{N-1},\\
&\frac{\partial f}{\partial \hat{t}_{y}} = -\frac{1}{N(N-1)}\Big(\hat{t}_z + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}}\Big),\\
&\frac{\partial f}{\partial \hat{t}_{z}} = -\frac{1}{N(N-1)}\Big(\hat{t}_y + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}\Big),\\
&\frac{\partial f}{\partial \hat{t}_{yzj}} = \frac{1}{N-1}(\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\boldsymbol{\tau}_j,\ j=1,2,\\
&\frac{\partial f}{\partial \hat{t}_{yj}} = -\frac{1}{N(N-1)}(\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\boldsymbol{\tau}_j\Big(\hat{t}_z + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}}\Big),\ j=1,2,\\
&\frac{\partial f}{\partial \hat{t}_{zj}} = -\frac{1}{N(N-1)}\Big(\hat{t}_y + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}\Big)(\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\boldsymbol{\tau}_j,\ j=1,2,\\
&\frac{\partial f}{\partial \hat{t}_{j}} = \frac{1}{N-1}\Big(-\boldsymbol{\tau}_j'\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{yz\mathbf{x}} + \frac{1}{N}\big(\hat{t}_y\boldsymbol{\tau}_j'\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}} + \hat{t}_z\boldsymbol{\tau}_j'\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}\\
&\qquad + \boldsymbol{\tau}_j'\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}(\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}} + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}}\boldsymbol{\tau}_j'\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}}\big)\Big),\ j=1,2,\\
&\frac{\partial f}{\partial \hat{t}_{ij}} = \frac{1}{N-1}\Big((\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')(-\widehat{\mathbf{A}}^{-1}\boldsymbol{\Lambda}_{ij}\widehat{\mathbf{A}}^{-1})\hat{\mathbf{t}}_{yz\mathbf{x}}\\
&\qquad - \frac{1}{N}\big((\hat{t}_y + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{y\mathbf{x}})(\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')(-\widehat{\mathbf{A}}^{-1}\boldsymbol{\Lambda}_{ij}\widehat{\mathbf{A}}^{-1})\hat{\mathbf{t}}_{z\mathbf{x}}\\
&\qquad + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')(-\widehat{\mathbf{A}}^{-1}\boldsymbol{\Lambda}_{ij}\widehat{\mathbf{A}}^{-1})\hat{\mathbf{t}}_{y\mathbf{x}}(\hat{t}_z + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}^{-1}\hat{\mathbf{t}}_{z\mathbf{x}})\Big), i \le j = 1,2,
\end{aligned}
$$

čia $\boldsymbol{\tau}_j$ – dvimatis vektorius, kurio j-oji komponentė lygi 1, o kita – lygi 0; $\boldsymbol{\Lambda}_{ij}$ yra (2×2)-osios eilės kvadratinė matrica, kurios (i,j) ir (j,i) pozicijose esantys elementai lygūs 1, o visur kitur – nuliui.

Nesunkiai galima parodyti, kad

$$
\begin{aligned}
&E\hat{t}_{yz} = t_{yz} = \sum_{k=1}^{N} y_k z_k, \quad E\hat{t}_y = t_y = \sum_{k=1}^{N} y_k, \quad E\hat{t}_z = t_z = \sum_{k=1}^{N} z_k,\\
&E\hat{\mathbf{t}}_{yz\mathbf{x}} = \mathbf{t}_{yz\mathbf{x}} = \sum_{k=1}^{N} q_k y_k z_k \mathbf{x}_k, \quad E\hat{\mathbf{t}}_{y\mathbf{x}} = \mathbf{t}_{y\mathbf{x}} = \sum_{k=1}^{N} q_k y_k \mathbf{x}_k,
\end{aligned}
$$

$$E\hat{\mathbf{t}}_{z\mathbf{x}}=\mathbf{t}_{z\mathbf{x}}=\sum_{k=1}^{N}q_kz_k\mathbf{x}_k,\quad E\hat{\mathbf{t}}_{\mathbf{x}}=\mathbf{t}_{\mathbf{x}}=\sum_{k=1}^{N}\mathbf{x}_k,$$
$$E\widehat{\mathbf{A}}=\mathbf{A}=\sum_{k=1}^{N}q_k\mathbf{x}_k\mathbf{x}_k^{'}.$$

Dėl šios priežasties tašką, kurio aplinkoje skleisime funkciją $\widehat{Cov}_0(y,z)$ Teiloro eilute, pasirinksime taip:

$$\big(\hat{t}_{yz},\hat{t}_y,\hat{t}_z,\hat{\mathbf{t}}_{yz\mathbf{x}},\hat{\mathbf{t}}_{y\mathbf{x}},\hat{\mathbf{t}}_{z\mathbf{x}},\hat{\mathbf{t}}_{\mathbf{x}},\widehat{\mathbf{A}}\big)=\big(t_{yz},t_y,t_z,\mathbf{t}_{yz\mathbf{x}},\mathbf{t}_{y\mathbf{x}},\mathbf{t}_{z\mathbf{x}},\mathbf{t}_{\mathbf{x}},\mathbf{A}\big).$$

Apskaičiavę gautų išvestinių reikšmes minėtame taške, ištiesiname funkciją $\widehat{Cov}_0(y,z)$:

$$\begin{aligned}\widehat{Cov}_{0L}(y,z)&=\frac{1}{N-1}\Big(t_{yz}-\frac{1}{N}t_yt_z+1\big(\hat{t}_{yz}-t_{yz}\big)-\frac{1}{N}t_z\big(\hat{t}_y-t_y\big)\\&-\frac{1}{N}t_y\big(\hat{t}_z-t_z\big)+\sum_{j=1}^{2}B_j\big(\hat{t}_j-t_j\big)\Big)\\&=\frac{1}{N-1}\Big(D+\sum_{k\in\mathbf{s}}d_ky_kz_k-\frac{1}{N}t_z\sum_{k\in\mathbf{s}}d_ky_k-\frac{1}{N}t_y\sum_{k\in\mathbf{s}}d_kz_k\\&+B_1\sum_{k\in\mathbf{s}}d_ka_k+B_2\sum_{k\in\mathbf{s}}d_kb_k\Big)=\frac{1}{N-1}\Big(D+\sum_{k\in\mathbf{s}}d_ke_k\Big),\end{aligned}$$

čia

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

$$D=\frac{1}{N}t_yt_z-B_1t_1-B_2t_2,$$

$$e_k=y_kz_k-(\mu_zy_k-B_1a_k)-(\mu_yz_k-B_2b_k).$$

Dabar jau galime skaičiuoti kalibruotojo kovariacijos įvertinio $\widehat{Cov}_{w1}^{(tot)}(y,z)$ apytikslę dispersiją:

$$AVar\big(\widehat{Cov}_{w1}^{(tot)}(y,z)\big) = Var\big(\widehat{Cov}_{0L}(y,z)\big) = \frac{1}{(N-1)^2}Var\Big(D + \sum_{k\in\mathbf{s}} d_k e_k\Big)$$
$$= \frac{1}{(N-1)^2}Var\Big(\sum_{k\in\mathbf{s}} d_k e_k\Big).$$

Norint gauti teiginyje nurodytą kovariacijos įvertinio apytikslės dispersijos išraišką, belieka pasinaudoti Horvico ir Tompsono sumos įvertinio kintamajam $yz - (\mu_z y - B_1 a) - (\mu_y z - B_2 b)$ *dispersijos išraiška.*

□

Kalibruotojo populiacijos kovariacijos įvertinio, kurio svoriai w_k randami naudojantis tiesiniu kalibravimu, apytikslė dispersija gaunama kaip išvada iš 3.6 teiginio.

3.1 išvada. *Kalibruotojo kovariacijos įvertinio*

$$\widehat{Cov}_{w1}^{(lin)}(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k\Big(y_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i y_i\Big)\Big(z_k - \frac{1}{N}\sum_{i\in\mathbf{s}} w_i z_i\Big),$$

kurio svoriai w_k *yra randami naudojantis* (3.11) *kalibravimo lygtimi ir minimizuojant atstumo funkciją* L_1*, apibrėžtą* (1.12) *lygybe, apytikslė dispersija yra*

$$AVar(\widehat{Cov}_{w1}^{(lin)}(y,z)) = \frac{1}{(N-1)^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}e_k^\star e_l^\star,$$
$$\text{čia } \ e_k^\star = y_k z_k - \mu_z y_k - \mu_y z_k + Bc_k,$$

$$t_y = \sum_{k=1}^{N} y_k, \quad t_z = \sum_{k=1}^{N} z_k,$$
$$B = t_{cc}^{-1}\Big(-t_{yzc} + \frac{1}{N}(t_y t_{zc} + t_z t_{yc})\Big),$$

$$t_{cc} = \sum_{k=1}^{N} q_k c_k^2, \quad t_{yzc} = \sum_{k=1}^{N} q_k y_k z_k c_k$$

$$t_{yc} = \sum_{k=1}^{N} q_k y_k c_k, \quad t_{zc} = \sum_{k=1}^{N} q_k z_k c_k, \quad c_k = (a_k - \mu_a)(b_k - \mu_b).$$

Įrodant suformuluotą 3.6 teiginio išvadą, vietoj kalibruotųjų (3.29) svorių reikia imti

$$w_k = d_k \biggl(1 + \Bigl(\sum_{i=1}^{N} c_i - \sum_{i\in\mathbf{s}} d_i c_i\Bigr)\Bigl(\sum_{i\in\mathbf{s}} d_i q_i c_i^2\Bigr)^{-1} q_k c_k\biggr).$$

3.2 pastaba. Dispersijoms $Var(\widehat{Cov}_{w1}^{(tot)}(y,z))$, $Var(\widehat{Cov}_{w1}^{(lin)}(y,z))$ vertinti siūlome šiuos įvertinius:

$$\widehat{Var}(\widehat{Cov}_{w1}^{(tot)}(y,z)) = \frac{1}{(N-1)^2} \sum_{k\in\mathbf{s}} \sum_{l\in\mathbf{s}} \Bigl(1 - \frac{\pi_k \pi_l}{\pi_{kl}}\Bigr) \frac{\hat{e}_k}{\pi_k} \frac{\hat{e}_l}{\pi_l},$$

$$\widehat{Var}(\widehat{Cov}_{w1}^{(lin)}(y,z)) = \frac{1}{(N-1)^2} \sum_{k\in\mathbf{s}} \sum_{l\in\mathbf{s}} \Bigl(1 - \frac{\pi_k \pi_l}{\pi_{kl}}\Bigr) \frac{\hat{e}_k^\star}{\pi_k} \frac{\hat{e}_l^\star}{\pi_l}.$$

Dydžiai $\hat{e}_k$, $\hat{e}_k^\star$ yra apibrėžiami pakeičiant atitinkamose išraiškose e_k ir $e_k^\star$ esančius nežinomus parametrus $\mathbf{A}$, $\mathbf{t}_{yz\mathbf{x}}$, t_y, $\mathbf{t}_{z\mathbf{x}}$, t_z, $\mathbf{t}_{y\mathbf{x}}$ ir t_{cc}, t_{yzc}, t_y, t_{zc}, t_z, t_{yc} jų įverčiais:

$$\widehat{\mathbf{A}} = \sum_{k\in\mathbf{s}} d_k q_k \mathbf{x}_k \mathbf{x}_k', \quad \hat{\mathbf{t}}_{yz\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k y_k z_k \mathbf{x}_k, \quad \hat{t}_y = \sum_{k\in\mathbf{s}} d_k y_k,$$
$$\hat{\mathbf{t}}_{z\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k z_k \mathbf{x}_k, \quad \hat{t}_z = \sum_{k\in\mathbf{s}} d_k z_k, \quad \hat{\mathbf{t}}_{y\mathbf{x}} = \sum_{k\in\mathbf{s}} d_k q_k y_k \mathbf{x}_k$$

ir

$$\hat{t}_{cc} = \sum_{k\in\mathbf{s}} d_k q_k c_k^2, \quad \hat{t}_{yzc} = \sum_{k\in\mathbf{s}} d_k q_k y_k z_k c_k, \quad \hat{t}_y = \sum_{k\in\mathbf{s}} d_k y_k,$$
$$\hat{t}_{zc} = \sum_{k\in\mathbf{s}} d_k q_k z_k c_k, \quad \hat{t}_z = \sum_{k\in\mathbf{s}} d_k z_k, \quad \hat{t}_{yc} = \sum_{k\in\mathbf{s}} d_k q_k y_k c_k.$$

3.3 pastaba. Tikslesni dispersijos įvertiniai gali būti sukonstruoti ir kelias svorių sistemas naudojantiems 3.2.4 skyrelio įvertiniams $\widehat{Cov}_{mw}^{(2)}$, $\widehat{Cov}_{mw}^{(4)}$. Šių įvertinių svorių sistemos apibrėžiamos tiesinėmis svorių w_k atžvilgiu kalibravimo lygtimis ir naudojant standartines atstumo funkcijas. Todėl nekyla problemų dėl svorių neišreikštinio pavidalo.

3.4. Kalibruotieji ir modeliais pagrįsti baigtinės populiacijos kovariacijos įvertiniai

Šiame poskyryje nagrinėsime dviejų tipų baigtinės populiacijos kovariacijos įvertinius. Pirmo tipo įvertinius jau aptarėme aukštesniuose 3.2 ir 3.3 poskyriuose. Tie įvertiniai yra sukonstruoti taikant įvairias kalibravimo lygtis ir atstumo funkcijas ir naudoja vieną arba kelias svorių sistemas. Iš šios plačios klasės įvertinių paimsime tokį įvertinį, kuris naudoja dvi svorių sistemas:

$$\widehat{Cov}_{mw}^{(4)}(y,z) = \frac{1}{N-1}\sum_{k\in\mathbf{s}} w_k^{[1]}\left(y_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[2]} y_l\right)\left(z_k - \frac{1}{N}\sum_{l\in\mathbf{s}} w_l^{[2]} z_l\right), \tag{3.30}$$

čia svoriai $w_k^{[1]}$ apibrėžti naudojantis tiesiniu kalibravimu ir atstumo funkcijomis L_1 bei L_6, o svoriai $w_k^{[2]}$ – naudojantis žinomų sumų kalibravimu bei tomis pačiomis atstumo funkcijomis. Kaip matome, šį įvertinį jau nagrinėjome 3.2.4 skyrelyje (4 atvejis), kalibruotiems svoriams apibrėžti naudodami funkciją L_1.

Antro tipo populiacijos kovariacijos įvertiniai yra sukonstruoti kalibruojant imties plano svorius ir naudojant tiesinį regresinį modelį. Šiuos įvertinius sukonstravo Sitter ir Wu [100]. Juos apibrėšime kitame skyrelyje. Kai kurie šio poskyrio rezultatai yra paskelbti [A11] publikacijoje.

3.4.1. Parametriniais modeliais pagrįsti kalibruotieji baigtinės populiacijos kovariacijos įvertiniai

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$, kurią sudaro N elementų. Kiekvienam populiacijos elementui u_k priskirkime M tyrimo ir $J+1$ papildomų kintamųjų reikšmių vektorius

$$\mathbf{y}_k = (y_{1k}, y_{2k}, \ldots, y_{Mk})' \text{ ir } \mathbf{x}_k = (1, x_{1k}, x_{2k}, \ldots, x_{Jk})'.$$

Sitter ir Wu [100] straipsnyje vertinamaisiais parametrais buvo pasirinktos kvadratinės baigtinės populiacijos funkcijos. Kiekviena kvadratinė funkcija gali būti užrašyta taip:

$$T(\mathbf{y}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \phi(\mathbf{y}_i, \mathbf{y}_j), \tag{3.31}$$

čia $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$, $\phi(\cdot, \cdot)$ yra simetrinė funkcija. Išrašius šios lygybės visus dvigubos sumos dėmenis, sutraukus panašius narius ir juos sunumeravus, funkcija T gali būti išreikšta ir taip:

$$T(\mathbf{y}) = \sum_{\alpha=1}^{N^*} \psi_\alpha,$$

čia α yra indeksų poros (ij) numeris sekoje visų galimų porų, tenkinančių sąlygą $i < j$; atitinkamą indeksų porų imtį pažymėkime $\mathbf{s}^* = \{\alpha = (ij) \mid \ i < j \ \text{ ir } i, j \in \mathbf{s}\}$; $\psi_\alpha = \phi(\mathbf{y}_i, \mathbf{y}_j)$, $N^* = N(N-1)/2$.

Taigi T yra populiacijos, turinčios N^* elementų, suma. Ji gali būti vertinama modeliais pagrįstais kalibruotaisiais populiacijos sumos įvertiniais, kurie buvo sukonstruoti tų pačių autorių (Sitter ir Wu) [115] straipsnyje. Buvo padaryta prielaida, kad tyrimo kintamuosius ir papildomus kintamuosius sieja pusiau parametrinis modelis. Jo analogas vieno tyrimo kintamojo atveju apibrėžiamas (1.27) lygybėmis. Padarę minėtą prielaidą, Sitter ir Wu imties duomenų pagrindu įvertino modelio parametrus, kuriuos vėliau panaudojo skaičiuodami $\mathbf{y}_i$ prognozuojamas reikšmes $\widehat{\mathbf{y}}_i$ ir apibrėždami modeliu pagrįstą ir kalibruotą baigtinės populiacijos kvadratinių funkcijų įvertinį

$$\widehat{T}_{MC} = \sum \sum_{(ij) \in \mathbf{s}^\star} w_{ij} \phi(\mathbf{y}_i, \mathbf{y}_j), \tag{3.32}$$

čia svoriai w_{ij} tenkina kalibravimo lygtis

$$\frac{1}{N^\star} \sum_{i \in \mathbf{s}} \sum_{j > i} w_{ij} = 1, \quad \sum_{i \in \mathbf{s}} \sum_{j > i} w_{ij} \phi(\widehat{\mathbf{y}}_i, \widehat{\mathbf{y}}_j) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \phi(\widehat{\mathbf{y}}_i, \widehat{\mathbf{y}}_j)$$

ir minimizuoja funkciją

$$\Phi = \sum_{i \in \mathbf{s}} \sum_{j > i} (w_{ij} - d_{ij})^2 / d_{ij} q_{ij}.$$

Svoriai q_{ij} yra teigiami ir laisvai pasirenkami, o $d_{ij} = 1/\pi_{ij}$.

Išsprendus kalibravimo uždavinį, (3.32) įvertinys įgyja šį pavidalą:

$$\widehat{T}_{MC} = \sum_{i \in \mathbf{s}} \sum_{j>i} d_{ij} \phi(\mathbf{y}_i, \mathbf{y}_j) + \Big\{ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \phi(\widehat{\mathbf{y}}_i, \widehat{\mathbf{y}}_j) - \sum_{i \in \mathbf{s}} \sum_{j>i} d_{ij} \phi(\widehat{\mathbf{y}}_i, \widehat{\mathbf{y}}_j) \Big\} \widehat{B}, \tag{3.33}$$

$$\widehat{B} = C(u, v)/C(u, u), \tag{3.34}$$

čia

$$C(u, v) = \sum \sum_{(ij) \in \mathbf{s}^*} d_{ij} q_{ij} (u_{ij} - \bar{u})(v_{ij} - \bar{v}),$$

$$u_{ij} = \phi(\widehat{\mathbf{y}}_i, \widehat{\mathbf{y}}_j), \quad v_{ij} = \phi(\mathbf{y}_i, \mathbf{y}_j),$$

$\bar{u} = \sum \sum_{(ij) \in \mathbf{s}^*} d_{ij} q_{ij} u_{ij} / \sum \sum_{(ij) \in \mathbf{s}^*} d_{ij} q_{ij}$; $\bar{v}$ ir $C(u, u)$ apibrėžti analogiškai.

Atskiru atveju, kai $\phi(\mathbf{y}_i, \mathbf{y}_j) = \frac{1}{N(N-1)} (y_i - y_j)(z_i - z_j)$, $\mathbf{y}_i = (y_i, z_i)'$, kvadratinė funkcija T yra populiacijoje $\mathcal{U}$ apibrėžtų kintamųjų y ir z kovariacija

$$Cov(y, z) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} (y_i - y_j)(z_i - z_j).$$

Toliau tarkime, kad y_i ir $\mathbf{x}_i$ (z_i ir $\mathbf{x}_i$) ryšius aprašo tiesinis regresinis modelis ξ:

$$E_\xi(y_i) = \mathbf{x}_i' \boldsymbol{\beta}, \quad E_\xi(z_i) = \mathbf{x}_i' \boldsymbol{\gamma}. \tag{3.35}$$

Tada iš bendrosios įvertinio išraiškos (3.33) kvadratinėms funkcijoms Sitter ir Wu gavo modeliu ξ pagrįstą ir kalibruotąjį populiacijos kovariacijos įvertinį

$$\widehat{Cov}_{MC}(y, z) = \widehat{Cov}_{HT} + \widehat{\boldsymbol{\beta}}' (S^2_{\mathbf{x}} - s^2_{\mathbf{x}}) \widehat{\boldsymbol{\gamma}} \widehat{B}, \tag{3.36}$$

čia $\widehat{\boldsymbol{\beta}} = \big\{ \sum_{i \in \mathbf{s}} d_i \mathbf{x}_i \mathbf{x}_i' \big\}^{-1} \sum_{i \in \mathbf{s}} d_i \mathbf{x}_i y_i$ yra imties planu pagrįstas regresijos koeficientų vektoriaus $\boldsymbol{\beta}$ įvertinys. Vektoriaus $\boldsymbol{\gamma}$ įvertinys yra apibrėžiamas analogiškai;

$$\widehat{Cov}_{HT} = \frac{1}{N(N-1)} \sum_{i \in \mathbf{s}} \sum_{j>i} d_{ij} (y_i - y_j)(z_i - z_j),$$

$$S^2_{\mathbf{x}} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{X}})(\mathbf{x}_i - \overline{\mathbf{X}})',$$

$$s_{\mathbf{x}}^2 = \frac{1}{N(N-1)} \sum_{i \in \mathbf{s}} \sum_{j>i} d_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'.$$

Kitame skyrelyje modeliuodami lyginsime modeliais pagrįstus ir kalibruotus baigtinės populiacijos kovariacijos įvertinius su kalibruotais baigtinės populiacijos kovariacijos (3.30) įvertiniais, kurie naudoja dvi svorių sistemas. Lyginant minėtus įvertinius teoriškai, iškyla tam tikrų sunkumų. Dispersijų išraiškos yra labai sudėtingos ir lyginant apytiksles dispersijas gaunamos komplikuotos nelygybės. Sunku nustatyti sąlygas, kurioms galiojant šios nelygybės yra teisingos.

3.4.2. Įvertinių tikslumo tyrimas, atliekant matematinį modeliavimą

Modeliuodami lyginsime du kalibruotuosius (3.30) pavidalo įvertinius su tiesiniu regresiniu modeliu pagrįstu kalibruotuoju baigtinės populiacijos kovariacijos (3.36) įvertiniu. Pirmųjų dviejų įvertinių svoriams $w_k^{[1]}$ ir $w_k^{[2]}$ apibrėžti naudotos atstumo funkcijos L_1 ir L_6. Šias atstumo funkcijas atitinkančius kalibruotuosius įvertinius pažymėjome $\widehat{Cov}_{mw1}^{(4)}(y,z)$, $\widehat{Cov}_{mw6}^{(4)}(y,z)$. Norėdami parodyti naudojamos papildomos informacijos privalumus, kalibruotuosius įvertinius lyginome su paprastu, tik imties planu pagrįstu kovariacijos įvertiniu $\widehat{Cov}_1(y,z)$.

Nagrinėjome tą pačią (žr. 3.2.3 skyrelį) 300 elementų dydžio realią populiaciją iš Lietuvos įmonių tyrimo. Ši populiacija buvo suskaidyta į du sluoksnius, atsižvelgiant į tyrimo kintamojo y didumą: populiacijos elementai, atitinkantys y reikšmes, mažesnes už kintamojo y vidurkį μ_y, sudaro pirmą sluoksnį, o likusieji elementai – antrąjį. Kaip ir ankstesniuose bandymuose, naudojome paprastąjį atsitiktinį sluoksninį ėmimą. Imties dydis $n = 100$ paskirstytas į sluoksnius, remiantis Neimano optimaliuoju paskirstymu.

Buvo išrinkta 1000 imčių ir kiekvienai iš jų įvertinta populiacijos kovariacija, naudojant du kalibruotus įvertinius $\widehat{Cov}_{mw1}^{(4)}(y,z)$, $\widehat{Cov}_{mw6}^{(4)}(y,z)$, tiesiniu regresiniu modeliu pagrįstą ir kalibruotą įvertinį $\widehat{Cov}_{MC}(y,z)$ bei imties planu pagrįstą ir papildomos informacijos nenaudojantį įvertinį $\widehat{Cov}_1(y,z)$. Įvairiems papildomiems kintamiesiems, skirtingai koreliuojantiems su tyrimo kintamaisiais, 3.9 lentelėje pateikta naudotų įvertinių empirinė dispersija, poslinkis, vidutinė kvadratinė paklaida ir variacijos koeficientas. Modeliu pagrįstam kalibruotam kovariacijos įvertiniui naudojome trimatį papildomų kintamųjų vektorių $\mathbf{x} = (1, a, b)'$ ir lygius papildomus svorius $q_{ij} = 1$. Vienodi papildomi svoriai $q_k = 1$ taip pat buvo naudojami ir kalibruotiems įvertiniams $\widehat{Cov}_{mw1}^{(4)}(y,z)$, $\widehat{Cov}_{mw6}^{(4)}(y,z)$.

Iš 3.9 lentelės matyti, kad kalibruotųjų kovariacijos įvertinių $\widehat{Cov}_{mw1}^{(4)}(y,z)$ ir

3.9 lentelė. *Pagrindinės baigtinės populiacijos kovariacijos įvertinių tikslumo charakteristikos (tikroji kovariacijos reikšmė $Cov(y,z) = 66\,083\,066$)*

Įvertinys	Įvertis	Dispersija $\times\ 10^{-13}$	Poslinkis	$VKP \times 10^{-13}$	cv
$\rho(y,a) = 0{,}81$	$\rho(z,b) = 0{,}90$	$\rho(y,b) = 0{,}63$	$\rho(z,a) = 0{,}60$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	66 187 636	2,206	104 570	2,207	0,071 0
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	66 064 504	2,176	-18 562	2,176	0,070 6
$\widehat{Cov}_{MC}(y,z)$	75 642 103	8,904	9 559 036	18,041	0,124 7
$\widehat{Cov}_1(y,z)$	61 292 369	9,924	-4 790 697	12,219	0,162 5
$\rho(y,a) = 0{,}21$	$\rho(z,b) = 0{,}90$	$\rho(y,b) = 0{,}63$	$\rho(z,a) = 0{,}15$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	60 549 920	10,043	-5 533 146	13,104	0,165 5
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	60 510 963	10,005	-5 572 103	13,110	0,165 3
$\widehat{Cov}_{MC}(y,z)$	83 499 670	24,440	17 416 604	54,774	0,187 2
$\widehat{Cov}_1(y,z)$	61 029 215	10,432	-5 053 851	12,986	0,167 4
$\rho(y,a) = 0{,}23$	$\rho(z,b) = 0{,}31$	$\rho(y,b) = 0{,}19$	$\rho(z,a) = 0{,}16$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	60 959 385	10,438	-5 123 681	13,064	0,167 6
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	60 953 712	10,423	-5 129 355	13,054	0,167 5
$\widehat{Cov}_{MC}(y,z)$	60 990 137	9,534	-5 092 930	12,128	0,160 1
$\widehat{Cov}_1(y,z)$	61 173 915	10,365	-4 909 151	12,775	0,166 4

$\widehat{Cov}^{(4)}_{mw6}(y,z)$ tikslumas yra labai panašus. Taip yra dėl to, kad abu įvertiniai priklauso tai pačiai įvertinių klasei ir jų konstravimui buvo naudotasi labai panašiomis atstumo funkcijomis.

Esant gerai koreliuotiems papildomiems kintamiesiems, įvertiniai, kurie naudoja dvi svorių sistemas, yra geriausi, nepaisant mūsų išankstinio spėjimo, kad modeliu pagrįsto kalibruoto kovariacijos įvertinio tikslumas yra panašus į įvertinių $\widehat{Cov}^{(4)}_{mw1}(y,z)$, $\widehat{Cov}^{(4)}_{mw6}(y,z)$ tikslumą ar net didesnis.

Kai turime tik vieną gerai koreliuotą su atitinkamu tyrimo kintamuoju papildomą kintamąjį, ($\rho(y,a) = 0{,}21$ ir $\rho(z,b) = 0{,}90$), tiesiniu regresiniu modeliu pagrįstas kalibruotasis įvertinys turi dižiausią dispersiją ir VKP, o įvertiniai $\widehat{Cov}^{(4)}_{mw1}(y,z)$, $\widehat{Cov}^{(4)}_{mw6}(y,z)$ yra truputį tikslesni už standartinį imties planu pagrįstą kovariacijos įvertinį $\widehat{Cov}_1(y,z)$.

Kai tyrimo ir atitinkamų papildomų kintamųjų koreliacija yra silpna, tai visi įvertiniai yra panašios kokybės. Modeliu pagrįstas kalibruotasis įvertinys turi šiek tiek mažesnę dispersiją ir VKP.

Baigiant šį skyrelį, yra sunku pasakyti, ar 3.9 lentelė pateiktų tą pačią informaciją eksperimentuojant su tokia populiacija, kurios elementų duomenys būtų idealiai aprašyti tiesiniu regresiniu modeliu. Tai paneigti ar patvirtinti, reikalingos tolesnės matematinio modeliavimo studijos.

3.4.3. Pataisytieji tiesiniu regresiniu modeliu pagrįsti ir kalibruoti baigtinės populiacijos kovariacijos įvertiniai

Baigtinės populiacijos kovariacijos įvertinius (3.36) galima modifikuoti (pataisyti). Juos konstruojant buvo tariama, kad ryšius tarp y_i ir $\mathbf{x}_i$ (z_i ir $\mathbf{x}_i$) aprašo tiesinis regresinis modelis: $E_\xi(y_i) = \mathbf{x}_i'\boldsymbol{\beta}$ ($E_\xi(z_i) = \mathbf{x}_i'\boldsymbol{\gamma}$). Taigi prognozuojamos reikšmės $\hat{y}_i$ ir $\hat{z}_i$ yra apskaičiuojamos naudojant tą patį trimatį papildomų kintamųjų vektorių. Čia ir slypi prastesnio kai kuriais atvejais vertinimo priežastis, nes galime susidurti su multikolinearumo problema, kai papildomi kintamieji koreliuoja tarpusavyje. Kintamųjų multikolinearumas mažina regresijos koeficientų stabilumą, o šie – įvertinių tikslumą. Mūsų buvusiame pavyzdyje, esant didelei papildomų ir tyrimo kintamųjų koreliacijai, koreliacija tarp papildomų kintamųjų yra lygi $\rho(a,b) = 0{,}54$. Tai jau gana didelė koreliacija.

Todėl mes siūlome naują idėją: nagrinėti šiuos regresinius modelius

$$E_{\xi_y}(y_i) = \beta_0 + \beta_1 a_i, \quad E_{\xi_z}(z_i) = \gamma_0 + \gamma_1 b_i. \tag{3.37}$$

Šiuo atveju kiekvieno tyrimo kintamojo prognozuojamos reikšmės apskaičiuojamos naudojant tik jo papildomą kintamąjį:

$$\hat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 a_i, \quad \hat{z}_i = \widehat{\gamma}_0 + \widehat{\gamma}_1 b_i, \tag{3.38}$$

čia

$$\widehat{\beta}_1 = \frac{\sum_{k\in\mathbf{s}} d_k \sum_{k\in\mathbf{s}} d_k a_k y_k - \sum_{k\in\mathbf{s}} d_k a_k \sum_{k\in\mathbf{s}} d_k y_k}{\sum_{k\in\mathbf{s}} d_k \sum_{k\in\mathbf{s}} d_k a_k^2 - \Big(\sum_{k\in\mathbf{s}} d_k a_k\Big)^2},$$

$$\widehat{\gamma}_1 = \frac{\sum_{k\in\mathbf{s}} d_k \sum_{k\in\mathbf{s}} d_k b_k z_k - \sum_{k\in\mathbf{s}} d_k b_k \sum_{k\in\mathbf{s}} d_k z_k}{\sum_{k\in\mathbf{s}} d_k \sum_{k\in\mathbf{s}} d_k b_k^2 - \Big(\sum_{k\in\mathbf{s}} d_k b_k\Big)^2},$$

$$\widehat{\beta}_0 = \frac{\sum_{k\in\mathbf{s}} d_k y_k}{\sum_{k\in\mathbf{s}} d_k} - \widehat{\beta}_1 \frac{\sum_{k\in\mathbf{s}} d_k a_k}{\sum_{k\in\mathbf{s}} d_k}, \quad \widehat{\gamma}_0 = \frac{\sum_{k\in\mathbf{s}} d_k z_k}{\sum_{k\in\mathbf{s}} d_k} - \widehat{\gamma}_1 \frac{\sum_{k\in\mathbf{s}} d_k b_k}{\sum_{k\in\mathbf{s}} d_k}.$$

Tada, esant didelei papildomų ir tyrimo kintamųjų koreliacijai, galima tikėtis gana tikslių tyrimo kintamųjų reikšmių prognozių ir tuo pačiu mažesnės modeliu pagrįsto kalibruoto įvertinio dispersijos.

Remiantis (3.37) modeliais iš (3.33) gaunamas *pataisytasis tiesiniu regresiniu* *modeliu pagrįstas kalibruotas baigtinės populiacijos kovariacijos įvertinys*:

$$\widehat{Cov}_{MC}^{(adj)}(y,z) = \widehat{Cov}_{HT} + \widehat{\beta}_1\big(Cov(a,b) - \widehat{Cov}(a,b)\big)\widehat{\gamma}_1\widehat{B}, \tag{3.39}$$

čia

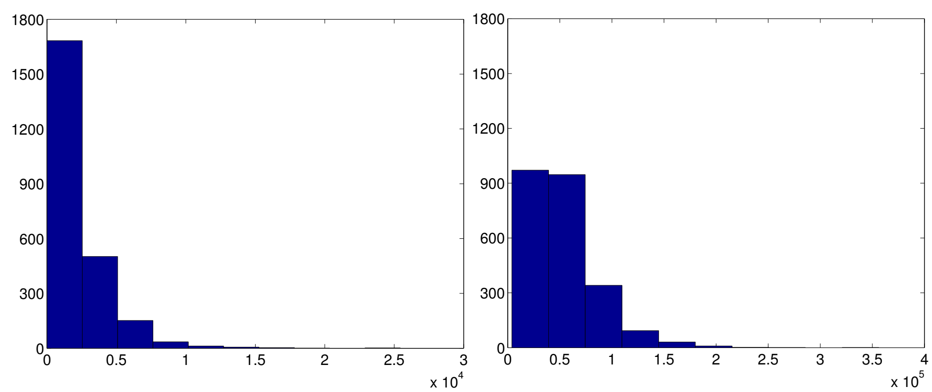
$$\widehat{Cov}(a,b) = \frac{1}{N(N-1)}\sum_{i\in\mathbf{s}}\sum_{j>i} d_{ij}(a_i - a_j)(b_i - b_j),$$

įvertinys $\widehat{B}$ apibrėžiamas (3.34) lygybe, kurioje tyrimo kintamųjų y ir z prognozuojamos reikšmės apskaičiuojamos naudojantis (3.38) formulėmis.

3.4.4. Pataisytojo tiesiniu regresiniu modeliu pagrįsto kovariacijos įvertinio savybių empirinis tyrimas

1 eksperimentas. Toliau dar kartą pakartosime 3.4.2 skyrelio eksperimentą, įtraukdami į jį naują populiacijos kovariacijos (3.39) įvertinį. Nauji rezultatai pateikiami 3.10 lentelėje.

3.4 pastaba. Visų įvertinių, išskyrus naująjį įvertinį, įvertintos tikslumo charakteristikos 3.10 lentelėje gali nežymiai skirtis nuo pateiktų 3.9 lentelėje rezultatų. Taip yra dėl to, kad kartojant bandymą buvo išrinktas kitas 1000 imčių rinkinys.



3.5 pav. *Wu ir Sitter naudota populiacija: tyrimo kintamųjų y (kairėje) ir z (dešinėje) reikšmių histogramos*

Esant didelei papildomų ir tyrimo kintamųjų koreliacijai, lygindami mode-

3.10 lentelė. *Baigtinės populiacijos kovariacijos įvertinių palyginimas (tikroji kovariacijos reikšmė $Cov(y,z) = 66\,083\,066$)*

Įvertinys	Įvertis	Dispersija $\times 10^{-13}$	Poslinkis	$VKP \times 10^{-13}$	cv
$\rho(y,a) = 0{,}81$	$\rho(z,b) = 0{,}90$	$\rho(y,b) = 0{,}63$	$\rho(z,a) = 0{,}60$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	66 161 336	2,224	78 269	2,225	0,071 3
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	66 036 227	2,194	-46 839	2,195	0,070 9
$\widehat{Cov}_{MC}(y,z)$	75 687 722	8,942	9 604 656	18,203	0,124 9
$\widehat{Cov}^{(adj)}_{MC}(y,z)$	65 661 651	2,151	-421 415	2,169	0,070 6
$\widehat{Cov}_{1}(y,z)$	60 997 364	9,879	-5 085 702	12,465	0,162 9
$\rho(y,a) = 0{,}21$	$\rho(z,b) = 0{,}90$	$\rho(y,b) = 0{,}63$	$\rho(z,a) = 0{,}15$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	60 655 410	9,855	-5 427 657	12,801	0,163 7
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	60 621 619	9,828	-5 461 447	12,811	0,163 5
$\widehat{Cov}_{MC}(y,z)$	83 402 177	24,317	17 319 111	54,355	0,186 9
$\widehat{Cov}^{(adj)}_{MC}(y,z)$	60 757 702	10,172	-5 325 364	13,008	0,166 0
$\widehat{Cov}_{1}(y,z)$	61 161 648	10,429	-4 921 418	12,851	0,167 0
$\rho(y,a) = 0{,}23$	$\rho(z,b) = 0{,}31$	$\rho(y,b) = 0{,}19$	$\rho(z,a) = 0{,}16$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	61 135 050	10,564	-4 948 017	13,012	0,168 1
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	61 125 941	10,556	-4 957 125	13,014	0,168 1
$\widehat{Cov}_{MC}(y,z)$	60 224 998	9,497	-5 858 068	12,980	0,161 8
$\widehat{Cov}^{(adj)}_{MC}(y,z)$	61 434 157	10,586	-4 648 910	12,747	0,167 5
$\widehat{Cov}_{1}(y,z)$	61 342 953	10,473	-4 740 113	12,720	0,166 8

liu pagrįstą kalibruotą kovariacijos įvertinį $\widehat{Cov}_{MC}(y,z)$ su naujuoju pataisytuoju įvertiniu $\widehat{Cov}^{(adj)}_{MC}(y,z)$ pastebime, kad pastarojo visos tikslumo charakteristikos yra kelis kartus mažesnės: dispersija sumažėjo maždaug 4 kartus, dėl kur kas mažesnio poslinkio VKP sumažėjo daugiau nei 8 kartus, variacijos koeficientas – beveik perpus.

Pažvelgus į antrąją 3.10 lentelės dalį, matyti, kad pataisytasis įvertinys vėl tikslesnis – jo dispersija mažesnė maždaug 2 kartus, o VKP – daugiau nei 4 kartus.

Tuo atveju, kai turime blogai su tyrimo kintamaisiais koreliuotus papildomus kintamuosius, visi įvertiniai yra panašios kokybės.

Reziumuojant galima teigti, kad idėja, naudoti du atskirus tiesinius regresinius (3.37) modelius, mūsų turimiems duomenims pasitvirtino.

2 eksperimentas. Straipsnio [100] autoriai Wu ir Sitter empiriškai lygino sukonstruotą modeliu pagrįstą kalibruotą populiacijos kovariacijos įvertinį su standartiniu Horvico ir Tompsono įvertiniu $\widehat{Cov}_{HT}$ naudodami realios populiacijos

3.11 lentelė. *Baigtinės populiacijos kovariacijos įvertinių palyginimas (tikroji kovariacijos reikšmė* $Cov(y,z) = 50\,190\,065$*)*

Įvertinys	Įvertis	Dispersija $\times\, 10^{-13}$	Poslinkis	$VKP \times 10^{-13}$	cv
$\rho(y,a) = 0{,}40$	$\rho(z,b) = 0{,}87$	$\rho(y,b) = 0{,}60$	$\rho(z,a) = 0{,}44$		
$\widehat{Cov}^{(4)}_{mw1}(y,z)$	50 000 836	3,431	-189 230	3,434	0,117 1
$\widehat{Cov}^{(4)}_{mw6}(y,z)$	50 000 331	3,430	-189 734	3,434	0,117 1
$\widehat{Cov}_{MC}(y,z)$	50 339 570	2,860	149 505	2,863	0,106 2
$\widehat{Cov}^{(adj)}_{MC}(y,z)$	50 068 025	3,459	-122 040	3,460	0,117 5
$\widehat{Cov}_1(y,z)$	50 091 548	3,920	-98 517	3,921	0,125 0

(žr. 3.5 pav.), turinčios $N = 2396$ elementus, duomenis. Tai – Ontario provincijoje vykdyto šeimų išlaidų tyrimo duomenys. Juos pavyko gauti iš pačių straipsnio autorių. Dabar empiriškai palyginsime mūsų nagrinėtus įvertinius naujoje populiacijoje.

Ankstesnį šio poskyrio eksperimentą mes pakartojome naudodami naujoje populiacijoje apibrėžtus šiuos tyrimo ir papildomus kintamuosius: y – metinės išlaidos drabužiams, z – bendros išlaidos, a – šeimos narių skaičius, b – šeimos pajamos. Modeliuojant buvo naudotos 500 elementų paprastosios atsitiktinės sluoksninės imtys. Matematinio modeliavimo rezultatai pateikti 3.11 lentelėje.

Yra tik vienas gerai koreliuotas papildomas kintamasis ($\rho(y,a) = 0{,}40$, $\rho(z,b) = 0{,}87$), todėl visų kalibruotųjų kovariacijos įvertinių tikslumo charakteristikos yra panašios į standartinio įvertinio $\widehat{Cov}_1(y,z)$ atitinkamas charakteristikas. Tarp įvertinių $\widehat{Cov}^{(4)}_{mw1}(y,z)$, $\widehat{Cov}^{(4)}_{mw6}(y,z)$ ir $\widehat{Cov}^{(adj)}_{MC}(y,z)$ pastebimi dar mažesni skirtumai. Tiesiniu regresiniu (3.35) modeliu pagrįstas įvertinys $\widehat{Cov}_{MC}(y,z)$ turi mažiausią dispersiją, vidutinę kvadratinę paklaidą ir mažiausią variacijos koeficientą. Iš čia galime daryti išvadą, kad naujos populiacijos duomenims aprašyti tinkamesnis yra (3.35) modelis, kuriame kiekvieno tyrimo kintamojo reikšmėms prognozuoti naudojami du tie patys papildomi kintamieji.

Taigi kurį iš dviejų kovariacijos įvertinių – $\widehat{Cov}_{MC}(y,z)$ ar $\widehat{Cov}^{(adj)}_{MC}(y,z)$ – pasirinkti tyrimui, priklauso nuo populiacijos duomenų.

3.5. Trečiojo skyriaus apibendrinimas

1. Taikant skirtingas atstumo funkcijas, sukonstruoti baigtinės populiacijos sumos kalibruotieji įvertiniai.

2. Išvestos kalibruotųjų svorių išraiškos baigtinės populiacijos kovariacijai vertinti. Vertinant šį parametrą vienu metu gali būti naudojamos net trys kalibruotųjų svorių sistemos. Šioms sistemoms apibrėžti pasiūlytos naujos kalibravimo lygtys.
3. Modeliuojant atlikta skirtingų kalibravimo lygčių įtakos populiacijos kovariacijos vertinimo tikslumui analizė.
4. Pasiūlyti sukonstruotų kalibruotųjų populiacijos sumos ir kovariacijos (dispersijos) įvertinių dispersijų vertinimo būdai.
5. Modifikuotas tiesiniu regresiniu modeliu pagrįstas kalibruotasis baigtinės populiacijos kovariacijos įvertinys.

Bendrosios išvados

Išsprendus įvade suformuluotas problemas, gauti šie rezultatai:

1. Išvestos kalibruotųjų svorių išraiškos baigtinės populiacijos sumai vertinti. Rezultatai gauti naudojant septynias skirtingas atstumo funkcijas. Sukonstruoti įvertiniai yra kur kas tikslesni už standartinius įvertinius, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė.

2. Taikant skirtingas atstumo funkcijas ir kalibravimo lygtis, sukonstruoti trijų tipų baigtinės populiacijos kovariacijos kalibruotieji įvertiniai, naudojantys atitinkamai vieną, dvi bei tris svorių sistemas. Matematinio modeliavimo rezultatai rodo, kad kalibruotieji kovariacijos įvertiniai yra tikslesni už standartinius įvertinius, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė. Svorių sistemų skaičiaus padidinimas iki 2 ar 3 daugeliu atvejų sumažina įvertinių dispersiją.

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

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

5. Sukonstruotas pataisytasis tiesiniu regresiniu modeliu pagrįstas kalibruotasis baigtinės populiacijos kovariacijos įvertinys. Modeliuojant pastebėta,

kad šio įvertinio vidutinė kvadratinė paklaida yra mažesnė už Wu ir Sitter [100] pasiūlyto įvertinio vidutinę kvadratinę paklaidą, jei tyrimo ir papildomų kintamųjų koreliacija yra didelė.

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

Tolesni tyrimo darbai sietini su šiomis temomis:

1. Sukonstruotų kalibruotųjų kovariacijos įvertinių dispersijų vertinimas taikant įtakos funkcijų ištiesinimo metodą.
2. Modeliais pagrįstų ir kalibruotųjų įvertinių konstravimas vertinant parametrus populiacijos srityse.
3. Populiacijos sluoksniavimas turint daugiau nei vieną tyrimo kintamąjį.

Literatūros sąrašas

[1] Alkaya, A.; Esin, A. Calibration estimator. G.U. Journal of Science, 18, No. 4, 2005, p. 591–601.

[2] Andersson, P.; Thorburn, D. An optimal calibration distance leading to the optimal regression estimator. Survey Methodology, 31, No. 1, 2005, p. 95–99.

[3] Bartkus, I. Kalibruotieji populiacijos sumos įvertiniai esant neatsakymams. Baigiamasis bakalauro darbas, VPU, Vilnius, 2007.

[4] Beaumont, J. Calibrated imputation in surveys under a quasi-model-assisted approach. Journal of the Royal Statistical Society, 67, No. 3, 2005, p. 445–458.

[5] Bethlehem, J. Reduction of nonresponse bias through regression estimation. Journal of Official Statistics, 4, No. 3, 1988, p. 251–260.

[6] Bühler, W.; Deutler, T. Optimal stratification and grouping by dynamic programming. Metrika, 22, 1975, p. 161–175.

[7] Breidt, F.; Opsomer, J. Local polynomial regression estimators in survey sampling. The Annals of Statistics, 28, No. 4, 2000, p. 1026–1053.

[8] Cassel, C.; Särndal, C.; Wretman, J. Some results on generalized difference estimation and generalized regression estimation for finite populations.

Biometrika, 63, No. 3, 1976, p. 615–620.

[9] Chadyšas, V.; Krapavickaitė, D. Investigation of accuracy of a calibrated estimator of a ratio by modelling. Nonlinear Analysis: Modelling and Control, 10, No. 4, 2005, p. 333–342.

[10] Chambers, R.; Skinner, C.; S.Wang. Intelligent calibration. Bulletin of the International Statistical Institute, 58, No. 2, 1999, p. 321–324.

[11] Chen, J.; Qin, J. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika, 80, No. 1, 1993, p. 107-116.

[12] Chen, J.; Sitter, R. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statistica Sinica, 9, 1999, p. 385–406.

[13] Chen, J.; Wu, C. Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. Statistica Sinica, 12, 2002, p. 1223–1239.

[14] Cochran, W. Sampling Techniques. John Wiley and Sons, New York, 1977.

[15] Dalenius, T. The problem of optimum stratification. Skandinavisk Aktuarietidskrift, No. 3-4, 1950, p. 203–213.

[16] Dalenius, T.; Hodges, J. The choice of stratification points. Skandinavisk Aktuarietidskrift, 1957, p. 198–203.

[17] Dalenius, T.; Hodges, J. Minimum variance stratification. Journal of the American Statistical Association, 54, No. 285, 1959, p. 88–101.

[18] Demnati, A.; Rao, J. Linearization variance estimators for survey data. Survey Methodology, 30, No. 1, 2004, p. 17–26.

[19] Deng, L.; Wu, C. Estimation of variance of the regression estimator. Journal of the American Statistical Association, 82, No. 398, 1987, p. 568–576.

[20] Deville, J.; Särndal, C.; Sautory, O. Generalized raking procedures in survey sampling. Journal of the American Statistical Association, 88, No. 423, 1993, p. 1013–1020.

[21] Deville, J.; Särndal, C. Calibration estimators in survey sampling. Journal of the American Statistical Association, 87, 1992, p. 376–382.

[22] Durbin, J. A note on the application of quenouille's method of bias reduction to the estimation of ratios. Biometrika, 46, No. 3/4, 1959, p. 477–480.

[23] Čekanavičius, V.; Murauskas, G. Statistika ir jos taikymai, II. TEV, Vilnius, 2002.

[24] Ekman, G. An approximation useful in univariate stratification. The Annals of Mathematical Statistics, 30, No. 1, 1959, p. 219–229.

[25] Estevao, V.; Patak, Z.; Särndal, C. Maximizing the use of auxiliary information for calibration and regression estimation. // *Proceedings of the Section on Survey Research Methods*, Alexandria, VA, 1994. American Statistical Association.

[26] Estevao, V.; Särndal, C. A new perspective on calibration estimators. // *Proceedings of the Section on Survey Research Methods,* Alexandria, VA, 2003. American Statistical Association.

[27] Estevao, V.; Särndal, C. A functional form approach to calibration. Journal of Official Statistics, 16, No. 4, 2000.

[28] Estevao, V.; Särndal, C. The ten cases of auxiliary information for calibration in two-phase sampling. Journal of Official Statistics, 18, No. 2, 2002, p. 233–255.

[29] Estevao, V.; Särndal, C. Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. Journal of Official Statistics, 20, No. 4, 2004, p. 645–669.

[30] Estevao, V.; Särndal, C. Survey estimates by calibration on complex auxiliary information. International Statistical Review, 74, No. 2, 2006, p. 127–147.

[31] Farrell, P.; Singh, S. Penalized chi square distance function in survey sampling. // *ASA Proceedings*, New York, USA, 2002. American Statistical Association (available on CD).

[32] Farrell, P.; Singh, S. Model-assisted higher-order calibration of estimators of variance. Australian & New Zealand Journal of Statistics, 47, No. 3, 2005, p. 375–383.

[33] Fuller, W. Regression estimation for survey samples. Survey Methodology, 28, No. 1, 2002, p. 5–23.

[34] Ghosh, S. Optimum stratification with two characters. The Annals of Mathematical Statistics, 34, No. 3, 1963, p. 866–872.

[35] Gunning, P.; Horgan, J. A new algorithm for the construction of stratum boundaries in skewed populations. Survey Methodology, 30, No. 2, 2004, p. 159–166.

[36] Gunning, P.; Horgan, J.; Keogh, G. Efficient Pareto stratification. Mathematical Proceedings of the Royal Irish Academy, 106A, No. 2, 2006, p. 131-138.

[37] Gunning, P.; Horgan, J.; Yancey, W. Geometric stratification of accounting data. Contaduría y Administración, No. 214, 2004.

[38] Harms, T. Extensions of the calibration approach. Calibration of distribution functions and its link to small area estimators. // *Chintex Working Paper*, Vol. 13. Federal Statistical Office, Germany, 2003.

[39] Harms, T.; Duchesne, P. On calibration estimation for quantiles. Survey Methodology, 52, No. 1, 2006, p. 37–52.

[40] Hartley, H.; Ross, A. Unbiased ratio estimators. Nature, 174, 1954, p. 270–271.

[41] Hedlin, D. On the stratification of highly skewed population. Report, Department of Social Statistics, University of Southampton, 1998.

[42] Hidiroglou, M. Double sampling. Survey Methodology, 27, No. 2, 2001, p. 143-154.

[43] Hidiroglou, M.; Särndal, C. Use of auxiliary information for two-phase sampling. Survey Methodology, 24, No. 1, 1998, p. 11–20.

[44] Hidiroglou, M.; Stukel, D.; Särndal, C. Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization. Survey Methodology, 22, No. 2, 1996, p. 117–125.

[45] Horgan, J. Stratification of skewed populations: A review. International Statistical Review, 74, No. 1, 2006, p. 67–76.

[46] Horvitz, D.; Thompson, D. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, No. 260, 1952, p. 663–685.

[47] Isaki, C. Variance estimation using auxiliary information. Journal of the American Statistical Association, 78, No. 381, 1983, p. 117–123.

[48] Isaki, C.; Fuller, W. Survey design under the regression superpopulation model. Journal of the American Statistical Association, 77, No. 377, 1982, p. 89–96.

[49] Kalton, G.; I.Flores-Cervantes. Weighting methods. Journal of Official Statistics, 19, No. 2, 2003, p. 81–97.

[50] Kima, J.; Sungur, E.; Heo, T. Calibration approach estimators in stratified

sampling. Statistics & Probability Letters, 77, No. 1, 2007, p. 99–103.

[51] Kott, P. A practical use for instrumental-variable calibration. Journal of Official Statistics, 19, No. 3, 2003, p. 265–272.

[52] Kott, P. Comment on Demnati and Rao: Linearization variance estimators for survey data. Survey Methodology, 30, No. 1, 2004, p. 27–28.

[53] Kott, P. Using calibration weighting to adjust for nonresponse and coverage errors. Survey Methodology, 32, No. 2, 2006, p. 133–142.

[54] Kozak, M.; Verma, M. Geometric versus optimization approach to stratification: A comparison of efficiency. Survey Methodology, 32, No. 2, 2006, p. 157–163.

[55] Krapavickaitė, D.; Plikusas, A. Estimation of a ratio in the finite population. Informatica, 16, No. 3, 2005, p. 347–364.

[56] Krapavickaitė, D.; Plikusas, A. Imčių teorijos pagrindai. Technika, Vilnius, 2005.

[57] Kuk, A.; Mak, T. Median estimation in the presence of auxiliary information. Journal of the Royal Statistical Society, 51, No. 2, 1989, p. 261–269.

[58] Lavallée, P.; Hidiroglou, M. On the stratification of skewed populations. Survey Methodology, 14, 1988, p. 33–43.

[59] Lehtonen, R.; Särndal, C.; Veijanen, A. The effect of model choice in estimation for domains, including small domains. Survey Methodology, 29, No. 1, 2003, p. 33–44.

[60] Lipeikienė, J. Matematika su kompiuteriu. Kompiuterinės matematinės sistemos DERIVE, MAPLE, MATLAB. Mokslo aidai, Vilnius, 2002.

[61] Lundström, S. *Calibration as a Standard Method for Treatment of Nonresponse* Ph.d. thesis, Stockholm University, 1997.

[62] Lundström, S.; Särndal, C. Calibration as standard method for treatment of nonresponse. Journal of Official Statistics, 15, No. 2, 1999, p. 305–327.

[63] Mickey, M. Some finite population unbiased ratio and regression estimators. Journal of the American Statistical Association, 54, No. 287, 1959, p. 594–612.

[64] Montanari, G. Post-sampling efficient QR-prediction in large-sample surveys. International Statistical Review, 55, 1987, p. 191–202.

[65] Montanari, G.; Ranalli, M. On calibration methods for design based finite

population inferences. // *Bulletin of the International Statistical Institute*, Vol. LX, p. 81–82. 2003.

[66] Montanari, G.; Ranalli, M. Nonparametric model calibration estimation in survey sampling. Journal of the American Statistical Association, 100, No. 472, 2005, p. 1429–1442.

[67] Mukhopadhyay, P. Model calibration estimation of a finite population variance. Pakistan Journal of Statistics, 22, No. 1, 2006, p. 1–10.

[68] Murthy, M. Product method of estimation. Sankhya : the Indian Journal of Statistics Series A, 26, 1964, p. 69–74.

[69] Pascual, J. Unbiased ratio estimators in stratified sampling. Journal of the American Statistical Association, 56, No. 293, 1961, p. 70–87.

[70] Pla, L. Determining stratum boundaries with multivariate real data. Biometrics, 47, No. 4, 1991, p. 1409–1422.

[71] Plikusas, A. Sampling methods in Lithuanian official statistics. Design and parameter estimation problems. // *Probability Theory and Mathematical Statistics. Proceedings of the Seventh Vilnius Conference*, Uthrecht, 1999. VSP.

[72] Plikusas, A. Calibrated estimators of the ratio. Lietuvos matematikos rinkinys, 41, spec. numeris, 2001, p. 457–462.

[73] Plikusas, A. Calibrated weights for the estimators of the ratio. Lithuanian Mathematical Journal, 43, 2003, p. 543–547.

[74] Plikusas, A. Non-linear calibration. // *Proceedings, Workshop on Survey Sampling*, Riga, Latvia, 2006. Central Statistical Bureau of Latvia.

[75] Rao, J. Estimating totals and distribution functions using auxiliary information at the estimation stage. Journal of Official Statistics, 10, No. 2, 1994, p. 153-165.

[76] Rao, J.; Kovar, J.; Mantel, H. On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika, 77, No. 2, 1990, p. 365–375.

[77] Ren, R. Estimation de la fonction de répartition et des fractiles d'une population finie. Actes des journées de méthodologie statistique, INSEE Méthodes, 1, No. 100, 2002, p. 263–289.

[78] Rizvi, S.; Gupta, J.; Bhargava, M. Effect of optimum stratification on sampling with varying probabilities under proportional allocation. Statistica, 4,

2004, p. 721–733.

[79] Robinson, P.; Särndal, C. Asymptotic properties of the generalized regression estimator in probability sampling. Sankhya : the Indian Journal of Statistics Series B, 45, 1983, p. 240–248.

[80] Rueda, M.; Arcos, A. Improving ratio-type quantile estimates in a finite population. Statistical Papers, 45, 2004, p. 231–248.

[81] Rueda, M.; Martínez, S.; Martínez, H.; Arcos, A. Estimation of the distribution function with calibration methods. Journal of Statistical Planning and Inference, 137, 2007, p. 435–448.

[82] Rueda, M.; Martínez-Puertas, S.; Martínez-Puertas, H.; Arcos, A. Calibration methods for estimating quantiles. Metrika, 66, 2007, p. 355–371.

[83] Särndal, C. Efficient estimators with simple variance in unequal probability sampling. Journal of the American Statistical Association, 91, No. 435, 1996, p. 1289–1300.

[84] Särndal, C. The calibration approach in survey theory and practice. Survey Methodology, 33, No. 2, 2007, p. 99–119.

[85] Särndal, C.; Swensson, B.; Wretman, J. Model Assisted Survey Sampling. Springer-Verlag, NewYork, 1992.

[86] Särndal, C.-E. Topics in uses of auxiliary information in surveys: The role of models, Nonresponse adjustment, Estimation for (small) domains. // *Proceedings, Second Baltic-Nordic Conference on Survey Sampling*, Finland, 2007. Statistics Finland.

[87] Särndal, C.-E.; Lundström, S. Estimation in Surveys with Nonresponse. John Wiley and Sons, England, 2005.

[88] Searls, D. The utilization of a known coefficient of variation in the estimation procedure. Journal of the American Statistical Association, 59, No. 308, 1964, p. 1225–1226.

[89] Serfling, R. Approximately optimal stratification. Journal of the American Statistical Association, 63, No. 324, 1968, p. 1298–1309.

[90] Shao, J.; Tu, D. The Jackknife and Boostrap. Springer-Verlag, New York, 1995.

[91] Singh, H.; Espejo, M. On linear regression and ratio-product estimation of a finite population mean. The Statistician, 51, No. 1, 2003, p. 59–67.

[92] Singh, M. Ratio cum product method of estimation. Metrika, 12, 1967, p. 34-42.

[93] Singh, P.; Srivastava, A. Sampling schemes providing unbiased regression estimators. Biometrika, 67, No. 1, 1980, p. 205–209.

[94] Singh, R. Approximately optimum stratification on the auxiliary variable. Journal of the American Statistical Association, 66, No. 336, 1971, p. 829–833.

[95] Singh, S. Generalized calibration approach for estimating variance in survey sampling. Annals of the Institute of Statistical Mathematics, 53, No. 2, 2001, p. 404–417.

[96] Singh, S.; Horn, S.; Chowdhury, S.; Yu, F. Calibration of the estimators of variance. Australian & New Zealand Journal of Statistics, 41, No. 2, 1999, p. 199–212.

[97] Singh, S.; Joarder, A. Estimation of finite population variance using random non-response in survey sampling. Metrika, 47, 1998, p. 241–249.

[98] Singh, S.; Singh, H.; Upadhyaya, L. Chain ratio and regression type estimators for median estimation in survey sampling. Statistical Papers, 48, 2006, p. 23–46.

[99] Sitter, R. Variance estimation for the regression estimator in two-phase sampling. Journal of the American Statistical Association, 92, No. 438, 1997, p. 780–787.

[100] Sitter, R.; Wu, C. Efficient estimation of quadratic finite population functions in the presence of auxiliary information. Journal of the American Statistical Association, 97, No. 458, 2002, p. 535–543.

[101] Skinner, C. Calibration weighting and non-sampling errors. Research in Official Statistics, 2, 1999, p. 33–43.

[102] Srivastava, S. A generalized estimator for the mean of a finite population using multi-auxiliary information. Journal of the American Statistical Association, 66, 1971, p. 404–407.

[103] Srivastava, S.; Jhaajj, H. A class of estimators of the population mean in survey sampling using auxiliary information. Biometrika, 68, No. 1, 1981, p. 341–343.

[104] Theberge, A. Extension of calibration estimators in survey sampling. Journal of the American Statistical Association, 94, No. 446, 1999, p. 635–644.

[105] Thomsen, L. A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation. Metrika, 23, 1976, p. 15–25.

[106] Tille, Y. Unbiased estimation by calibration on distribution in simple sampling designs without replacement. Survey Methodology, 28, No. 1, 2002, p. 77-85.

[107] Tracy, D.; Singh, S.; Arnab, R. Note on calibration in stratified and double sampling. Survey Methodology, 29, No. 1, 2003, p. 99–104.

[108] Tukey, J. Bias and confidence in not-quite large samples. Annals of Mathematical Statistics, 29, No. 614, 1958.

[109] Verma, M. Optimum stratification for two sensitive quantitative variables using equal allocation method. Revista de Matemátika e Estatístika, 25, No. 2, 2007, p. 73–85.

[110] Welsh, A.; Ronchetti, E. Bias-calibrated estimation from sample surveys containing outliers. Journal of the Royal Statistical Society, 60, No. 2, 1998, p. 413–428.

[111] Winkler, W. Strata boundary determination. Statistical research division report, U.S. Bureau of the Census, 1998.

[112] Wu, C. Optimal calibration and empirical likelihood methods in survey sampling. // *Proceedings of Statistics Canada Symposium 2002: Modelling Survey Data for Social and Economic Research*, 2002.

[113] Wu, C. Optimal calibration estimators in survey sampling. Biometrika, 90, No. 4, 2003, p. 937–951.

[114] Wu, C.; Luan, Y. Optimal calibration estimators under two-phase sampling. Journal of Official Statistics, 19, No. 2, 2003, p. 119–131.

[115] Wu, C.; Sitter, R. A model-calibration approach to using complete auxiliary information from survey data. Journal of the American Statistical Association, 96, No. 453, 2001, p. 185–193.

[116] Yates, F.; Grundy, P. Selection without replacement from within strata with probability proportional to size. Journal of the Royal Statistical Society, 15, No. 2, 1953, p. 253–261.

Autoriaus publikacijų disertacijos tema sąrašas

Straipsniai recenzuojamuose periodiniuose mokslo leidiniuose

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

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

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

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

Straipsniai kituose mokslo leidiniuose

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

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

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

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

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

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

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

Sąvokų žodynas

A

apibendrintasis regresinis įvertinys – generalized regression estimator
apytikslė dispersija – approximate variance
asimetrijos koeficientas – skewness
asimetrinė populiacija – skewed population
asimptotinė dispersija – asymptotic variance
atsakymo tikimybė – response probability
atsitiktinis dydis – random variable
atstumo funkcija – distance function

B

baigtinė populiacija – finite population

D

dalinė išvestinė – partial derivative
dirbtinė populiacija – artificial population
diskretusis kintamasis – discrete variable
dispersija – variance
duomenys – data
dviejų fazių ėmimas – two-phase sampling

E

eksponentinis skirstinys – exponential distribution

F

funkcinės formos metodas – functional form approach

G

geometrinė progresija – geometric progression
geometrinis sluoksniavimo metodas – geometric stratification method

I

imčių tyrimas – sample survey
imties dydis – sample size
imties išrinkimas – sample selection
imties paskirstymas – allocation of a sample
imties planas – sample design
imties plano svoris – design weight
imties vidurkis – sample mean
imtis – sample
instrumentinis vektorius – instrument vector
išrinkimas – selection
iteracinė lygtis – iterative equation
ištiesintasis įvertinys – linearized estimator
išvestinė – derivative

Į

įrašymas – imputation
įvertinys – estimator
įvertis – estimate

K

kalibravimo lygtis – calibration equation
kalibruotasis svoris – calibrated weight
kalibruotasis įvertinys – calibrated estimator
kategorinis kintamasis – categorical variable
kiekybinis kintamasis – quantitative variable
koreliacija – correlation
koreliacijos koeficientas – correlation coefficient
kovariacija – covariance
kvadratinė funkcija – quadratic function
kvantilis – quantile

L

laipsnių metodas – power method

M

matematinė viltis – expected value
matematinis modeliavimas – mathematical simulation
mediana – median
minimaliosios lygtys – minimal equations
multikolinearumas – multicolinearity

N

neatsakymas – nonresponse
nepaslinktasis įvertinys – unbiased estimator
netiesinis kalibravimas – non-linear calibration
normalusis skirstinys – normal distribution

O

optimaliosios sluoksnių ribos – optimal stratum boundaries
optimalusis paskirstymas – optimal allocation

P

pagrįstas imties planu – design based
papildoma informacija – auxiliary information
papildomas kintamasis – auxiliary variable
paprastoji atsitiktinė imtis – simple random sample
parametras – parameter
pasiskirstymo funkcija – distribution function
pasitelkiantis modelį – model assisted
pataisytasis geometrinis sluoksniavimo metodas – adjusted geometric stratification method
pataisytasis modeliu pagrįstas įvertinys – adjusted model-calibrated estimator
populiacija – population
populiacijos parametras – population parameter
populiacijos sritis – population domain
poslinkis – bias
priklausymo tikimybė – inclusion probability
pseudoreikšmė – pseudovalue

R

realioji populiacija – real population
regresijos koeficientas – regression coefficient
regresinis įvertinys – regression estimator

S

santykinis įvertinys – ratio estimator
santykio įvertinys – estimator of the ratio
saviranka – bootstrap
skirstinys – distribution
sluoksnių ribos – stratum boundaries
sluoksniavimas – stratification
sluoksninis ėmimas – stratified sampling
sluoksnis – stratum
statistika – statistics
suderintasis įvertinys – consistent estimator
sumų kalibravimas – calibration of totals
suma – total

Š

šaknies iš f sluoksniavimo taisyklė – cumulative root frequency stratification method

T

tankis – density
Teiloro ištiesinimas – Taylor linearization
tiesinis kalibravimas – linear calibration
tiesinis regresinis modelis – linear regression model
tikimybė – probability
tikimybinė imtis – probability sample
tikroji reikšmė – true value
tikslumas – accuracy
tolydusis kintamasis – continuous variable
tyrimo kintamasis – study variable
tyrimo populiacija – survey population

V

variacijos koeficientas – coefficient of variation
vektorius – vector
vidurkis – mean
vidutinė kvadratinė paklaida – mean square error
visraktis – jackknife

Dalykinė rodyklė

Dalius Pumputis

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI,
GAUTI NAUDOJANT PAPILDOMĄ INFORMACIJĄ
Daktaro disertacija
Fiziniai mokslai, matematika (01P)

STATISTICAL ESTIMATORS OF THE FINITE POPULATION PARAMETERS
IN THE PRESENCE OF AUXILIARY INFORMATION
Doctoral Dissertation
Physical Sciences, Mathematics (01P)

2008 12 29. 8,5 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto leidykla „Technika“, Saulėtekio al. 11,
10223 Vilnius *http://leidykla.vgtu.lt*
Spausdino UAB „Baltijos kopija“, Kareivių g. 13B, 09109 Vilnius, *http://www.kopija.lt*