VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Viktoras CHADYŠAS

BAIGTINĖS POPULIACIJOS PARAMETRŲ STATISTINIAI ĮVERTINIAI ESANT IMTIES ROTACIJAI

DAKTARO DISERTACIJA

FIZINIAI MOKSLAI,
MATEMATIKA (01P)

Vilnius LEIDYKLA TECHNIKA 2009

Disertacija rengta 2005–2009 metais Vilniaus Gedimino technikos universitete.

Mokslinis vadovas
prof. habil. dr. Leonas SAULIS (Vilniaus Gedimino technikos universitetas, fiziniai mokslai, matematika – 01P).

Konsultantas
doc. dr. Danutė KRAPAVICKAITĖ (Matematikos ir informatikos institutas, fiziniai mokslai, matematika – 01P).

VGTU leidyklos TECHNIKA 1702-M mokslo literatūros knyga
http://leidykla.vgtu.lt

ISBN 978-9955-28-524-3
© VGTU leidykla TECHNIKA, 2009
© Chadyšas, V. 2009
viktoras.chadysas@fm.vgtu.lt

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Viktoras CHADYŠAS

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

DOCTORAL DISSERTATION

PHYSICAL SCIENCES,
MATHEMATICS (01P)

Vilnius LEIDYKLA TECHNIKA 2009

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2005–2009.

Scientific supervisor
Prof Dr Habil Leonas SAULIS (Vilnius Gediminas Technical University, Physical Sciences, Mathematics – 01P).

Consultant
Assoc Prof Dr Danutė KRAPAVICKAITĖ (Institute of Mathematics and Informatics, Physical Sciences, Mathematics – 01P).

Santrauka

Disertacijoje sudaromi baigtinės populiacijos tyrimo kintamojo sumos, pasiskirstymo funkcijos, kvantilio įvertiniai esant imties rotacijai.

Pirmiausia darbe nagrinėjamas baigtinės populiacijos tyrimo kintamojo sumos vertinimas esant imties rotacijai. Sudarytas sudėtinis santykinis sumos įvertinys naudojantis papildomą informaciją žinomą iš ankstesnių imties rinkimų. Modeliavimo rezultatai rodo, kad papildomos informacijos panaudojimas iš jau išrinktos imties gali pagerinti įvertinių tikslumą. Kelių ėmimų schema gali būti taikoma siekiant pagerinti baigtinės populiacijos nepaslinktojo imties planu pagrįsto sumos įvertinio tikslumą.

Taip pat nagrinėjami baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertiniai esant imties rotacijai. Sudaryti keli tyrimo kintamojo baigtinėje populiacijoje pasiskirstymo funkcijos sudėtiniai įvertiniai (regresinis ir santykinis) naudojant dviejų ėmimų schemą. Pasiūlyti optimalūs sudėtiniai pasiskirstymo funkcijos įvertiniai su mažiausia dispersija. Įvertiniai lyginami tarpusavyje atliekant modeliavimą su realiais duomenimis.

Baigtinės populiacijos kvantilio įvertiniai sudaromi imant sudėtinių pasiskirstymo funkcijų įvertinių atvirkštinių funkcijų įvertinius. Taip pat sudaromi pasikliautinojo intervalo įvertiniai kvantiliams, taikant kartotinių imčių metodus. Modeliuojant duomenis lyginamas jų tikslumas, daromos išvados apie pasikliautinojo intervalo įvertinių, sudarytų skirtingais būdais, kvantiliui efektyvumą.

Disertaciją sudaro įvadas, trys pagrindiniai skyriai, išvados ir literatūros sąrašas.

Tyrimų rezultatai paskelbti iš viso 6-iuose moksliniuose straipsniuose iš kurių keturi – recenzuojamuose mokslo leidiniuose. Šie rezultatai taip pat pristatyti 12-oje konferencijų.

Abstract

The dissertation analyzes how to incorporate auxiliary information into the estimation of the finite population total, distribution function and quantile in the case of sample rotation.

First of all estimation of the finite population total in the case of sample rotation are considered. We focus on construction of the total estimator for rotated sampling design. Successive sampling procedure using multi-phase sampling design have been developed. The composite ratio type estimator of the total using auxiliary information and its approximate variance is constructed. A simulation study, based on the real population data, is performed and the proposed estimators are compared by a traditional estimator for a total.

The composite estimators of finite population distribution function, constructed under sampling on two occasions, are considered. Composite regression and ratio type estimators are constucted, using values of the study variable as auxiliary information obtained on the first occasion. The optimal estimators, in the sense of minimal variance, is also obtained. A simulation study, based on the real population data, is performed and the proposed estimators are compared by a traditional estimator for a distribution function.

Several quantile estimators by deriving the distribution function estimators with the use of auxiliary information are proposed. Some procedures that may be used to obtain estimates of confidence intervals for quantiles in a finite population (most of which are based on resampling) are compared. A simulation study, based on two different artificial populations, is performed and comparisons of the estimation methods proposed for confidence intervals of population quantiles are made.

The thesis layout consists of introduction chapter, three main chapters, conclusion chapter and bibliography chapter.

The results of thesis were published in six scientific publications, four of which in the reviewed scientific publications. The results of thesis were announced in twelve conferences also.

Žymėjimai

Simboliai

$\mathcal{U}$	–	baigtinė populiacija, sudaryta iš N elementų
$\mathfrak{s}$	–	tikimybinė imtis
N	–	populiacijos dydis
n	–	imties dydis
y	–	tyrimo kintamasis
$p(\mathfrak{s})$	–	imties $\mathfrak{s}$ išrinkimo tikimybė
π_i	–	i-ojo populiacijos elemento priklausymo imčiai tikimybė
π_{ij}	–	elementų i ir j poros priklausymo imčiai tikimybė
t_y	–	kintamojo y suma
s_y^2	–	kintamojo y dispersija
μ_y	–	kintamojo y vidurkis
$\overline{y}$	–	kintamojo y imties vidurkis
$F_y(z)$	–	kintamojo y pasiskirstymo funkcija
K_{yq}	–	kintamojo y q lygmens kvantilis
M_y	–	kintamojo y mediana
R	–	dviejų populiacijos sumų t_y ir t_x santykis
$\rho(y,x)$	–	kintamųjų y ir x koreliacijos koeficientas

$\mathrm{Cov}(y,x)$	–	kintamųjų y ir x kovariacija
$\widehat{\theta}$	–	parametro θ įvertinys
$\mathrm{Posl}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ poslinkis
$\mathrm{SPosl}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ santykinis poslinkis
$\mathrm{D}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ dispersija
$\mathrm{AD}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ apytikslė dispersija
$\mathrm{VKP}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ vidutinė kvadratinė paklaida
$\mathrm{SVKP}(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ santykininė vidutinė kvadratinė paklaida
$cv(\widehat{\theta})$	–	įvertinio $\widehat{\theta}$ variacijos koeficientas

Turinys

ĮVADAS . . . 1
Tiriamoji problema . . . 1
Darbo aktualumas . . . 1
Tyrimų objektas . . . 2
Darbo tikslas ir uždaviniai . . . 2
Tyrimų metodai . . . 3
Mokslinis darbo naujumas . . . 3
Darbo rezultatų praktinė reikšmė . . . 3
Ginamieji disertacijos teiginiai . . . 3
Darbo rezultatų aprobavimas . . . 4
Autoriaus dalyvavimas mokslinėse programose . . . 5
Disertacijos struktūra . . . 6
Padėka . . . 7

1. Įvairių baigtinės populiacijos parametrų įvertiniai . . . 9
1.1. Žymenys ir pagrindinės sąvokos . . . 9
1.2. Įvertinių tikslumo matai . . . 11
1.3. Horvico ir Tompsono sumos įvertinys . . . 12
1.4. Sumos įvertiniai, naudojantys papildomą informaciją . . . 13
1.4.1. Santykinis įvertinys . . . 13
1.4.2. Regresinis įvertinys . . . 14

1.5. Imties planai 15
1.5.1. Paprastoji atsitiktinė imtis 15
1.5.2. Dviejų fazių ėmimas 17
1.6. Imties rotacija 18
1.7. Baigtinės populiacijos sudėtingesnių parametrų vertinimas 21
1.7.1. Imties planu pagrįstas pasiskirstymo funkcijos įvertinys . . . 21
1.7.2. Papildomą informaciją naudojantys pasiskirstymo funkcijos įvertiniai 22
1.7.3. Imties planu pagrįsti kvantilių įvertiniai 23
1.7.4. Papildomą informaciją naudojantys kvantilio įvertiniai 24
1.8. Sudėtingesnių įvertinių dispersijų vertinimas 27
1.8.1. Apytikslė įvertinio dispersija ir Teiloro ištiesinimo metodas jai rasti 27
1.8.2. Visrakčio metodas 28
1.8.3. Savirankos metodas 29
1.9. Kvantilio pasikliautinasis intervalas ir jo įvertiniai 30
1.10. Pirmojo skyriaus apibendrinimas. Uždavinių formulavimas 33

2. Baigtinės populiacijos sumos vertinimas esant imties rotacijai 35
2.1. Sumos vertinimas naudojant dviejų ėmimų schemą 36
2.1.1. Atskiras atvejis: paprastoji atsitiktinė negrąžintinė imtis . . . 42
2.1.2. Įvertinių tikslumo tyrimas, atliekant matematinį modeliavimą 46
2.2. Sumos vertinimas naudojant sudėtingesnę imties rinkimo schemą 48
2.2.1. Atskiras atvejis: paprastoji atsitiktinė negrąžintinė imtis . . 55
2.2.2. Matematinis modeliavimas: papildomos informacijos įtaka įvertinių tikslumui 58
2.3. Antrojo skyriaus apibendrinimas 59

3. Baigtinės populiacijos sudėtingesnių parametrų vertinimas esant imties rotacijai 61
3.1. Pasiskirstymo funkcijos vertinimas 61
3.1.1. Tradicinis pasiskirstymo funkcijos įvertinys 61
3.1.2. Regresinis pasiskirtymo funkcijos įvertinys ir jo apytikslė dispersija 64
3.1.3. Santykinis pasiskirstymo funkcijos įvertinys ir jo apytikslė dispersija 69
3.1.4. Įvertinių tikslumo tyrimas atliekant matematinį modeliavimą 72

3.2. Tiesioginiai kvantilių įvertiniai 77
3.2.1. Įvertinių empirinis palyginimas 79
3.3. Kvantilio pasikliautinojo intervalo vertinimas 81
3.3.1. Įvertinių tikslumo tyrimas atliekant matematinį modeliavimą 83
3.4. Trečiojo skyriaus apibendrinimas 87

Bendrosios išvados 89

Literatūros sąrašas 91

Autoriaus publikacijos disertacijos tema 95

Contents

Introduction 1
Scientific problem 1
Topicality of the work 1
Research object 2
The aim and tasks of the work 2
Applied methods 3
Scientific novelty 3
Practical value of the work results 3
Statements presented for defence 3
Approval of the work results 4
The scope of the scientific work 6
Acknowledgment 7

1. Various estimators of the finite population parameters 9
1.1. Definitions and notations 9
1.2. The specification of precision 11
1.3. Horvitz-Thompson estimator of the total 12
1.4. Estimators of the total using auxiliary information 13
1.4.1. Ratio estimator 13
1.2.2. Regression estimator 14
1.5. Sampling designs 15

"VikChaThesis" — 2009/12/21 — 13:05 — page xiv — #14

1.5.1. Simple random sampling 15
1.5.2. Two phase sampling 17
1.6. Sample rotation 18
1.7. Estimation of more complex parameters 21
1.7.1. Design based estimator of the distribution function 21
1.7.2. Estimators of the distribution function using auxiliary information 22
1.7.3. Design based estimators of the quantiles 23
1.7.4. Estimators of the quantiles using auxiliary information . . . 24
1.8. Variance estimation for complex estimators 27
1.8.1. An approximate variance of the estimator and Linearization (Taylor series) method to find it 27
1.8.2. Jackknife method 28
1.8.3. Bootstrap method 29
1.9. Confidence interval for quantile 30
1.10. Conclusions for Chapter 1 and revise of the tasks 33

2. Estimation of the total in the case of sample rotation 35
2.1. Estimation of the total using two occasion sampling scheme 36
2.1.1. A particular case: simple random sampling 42
2.1.2. Simulation study 46
2.2. Estimation of the total using more complex sampling scheme . . . 48
2.2.1. A particular case: simple random sampling 55
2.2.2. Simulation study 58
2.3. Conclusions for Chapter 2 59

3. More complex parameters estimation in the case of sample rotation 61
3.1. Estimation of the distribution function 61
3.1.1. Traditional estimator 61
3.1.2. Regression type estimator and its approximate variance . . 64
3.1.3. Ratio type estimator and its approximate variance 69
3.1.3. Simulation study 72
3.2. Direct quantile estimators 77
3.2.1. Simulation study 79
3.3. Estimation of the confidence interval for quantile 81
3.3.1. Simulation study 83
3.4. Conclusions for Chapter 3 87

General conclusions 89

References 91

List of the author's scientific publications on the topic of dissertation 95

Įvadas

Tiriamoji problema

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

Darbo aktualumas

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

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

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

Tyrimų objektas

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

Darbo tikslas ir uždaviniai

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

Siekiant numatyto tikslo buvo sprendžiami šie uždaviniai:

1. Sudaryti baigtinės populiacijos tyrimo kintamojo sumos įvertinius, naudojant papildomą informaciją, bei jų dispersijų įvertinius, naudojant imties rotaciją. Remiantis matematiniu modeliavimu ir naudojant realius duomenis palyginti siūlomus įvertinius su imties planu pagrįstu įvertiniu.
2. Sudaryti baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertinius bei pasiskirstymo funkcijos įvertinių atitinkamus dispersijos įvertinius, naudojant imties rotaciją. Sudaryti optimalius įvertinius, minimizuojančius pasiskirstymo funkcijų įvertinių dispersijas. Naudojant realius duomenis palyginti imties planu pagrįstą pasiskirstymo funkcijos įvertinį su autoriaus sukonstruotais pasiskirstymo funkcijos įvertiniais.
3. Vertinti baigtinės populiacijos tyrimo kintamojo kvantilį, naudojant šiame darbe pasiūlytų pasiskirstymo funkcijų įvertinių atvirkštinių funkcijų įvertinius. Pasiūlyti kelis kvantilių pasikliautinojo intervalo įvertinius, taikant kartotinių imčių metodus. Modeliuojant duomenis palyginti jų tikslumą.

Tyrimų metodai

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

Darbo mokslinis naujumas

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

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

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

Darbo rezultatų praktinė reikšmė

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

Ginamieji disertacijos teiginiai

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

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

Darbo rezultatų aprobavimas

Disertacijos tema paskelbtos 6 mokslinės publikacijos iš jų keturios recenzuojamuose mokslo leidiniuose. Disertacijos rezultatai aptarti šiose mokslinėse konferencijose ir seminaruose:

1. V. Chadyšas, Pasikliautinojo intervalo vertinimas kvantiliams baigtinėje populiacijoje, *10-oji Lietuvos jaunųjų mokslininkų konferencija*, VGTU, Vilnius, 2007.
2. V. Chadyšas, Estimation of the confidence intervals for quantiles in finite population, *12-th Mathematical Modelling and Analysis Conference*, Trakai, Lithuania, 2007.
3. V. Chadyšas, Estimation of the confidence intervals for quantiles in finite population, *Second Baltic-Nordic Conference on Survey Sampling*, Kuusamo, Finland, 2007.
4. V. Chadyšas, Pasikliautinojo intervalo vertinimas kvantiliams baigtinėje populiacijoje, *Lietuvos matematikų draugijos XLVIII-oji konferencija*, VGTU, Vilnius, 2007.
5. V. Chadyšas, Estimation of quantile in survey sampling, *Seminar in Survey Sampling*, University of Stocholm, Stocholm, Sweden, 2008.
6. V. Chadyšas, Kvantilio vertinimas baigtinėje populiacijoje, *11-oji respublikinė jaunųjų mokslinkų konferencija*, VGTU, Vilnius, 2008.

7. V. Chadyšas, Estimating quantiles under sampling on two occasions, *22nd Nordic Conference on Mathematical Statistics*, Vilnius, 2008.
8. V. Chadyšas, Baigtinės populiacijos pasiskirstymo funkcijos įvertinių konstravimas, naudojant imties rotaciją, *Lietuvos matematikų draugijos XLIX konferencija*, VDU, Kaunas, 2008.
9. D. Krapavickaitė, V. Chadyšas, Dirbančiųjų gyventojų skaičiaus vertinimas esant imties rotacijai, *Lietuvos matematikų draugijos XLIX konferencija*, VDU, Kaunas, 2008.
10. V. Chadyšas, Estimators of quantile for rotated sample design, *Workshop on Survey Sampling Theory and Methodology*, Kuressaare, Estonia, 2008.
11. V. Chadyšas, Papildomos informacijos panaudojimas konstruojant santykinį sumos įvertinį esant imties rotacijai, *Lietuvos matematikų draugijos L konferencija*, MII, Vilnius, 2009.
12. V. Chadyšas, Estimation of of total using auxiliary information, *The Baltic-Nordic-Ukrainian Summer School on Survey Statistics*, Kiyv, Ukraine, 2009.

Autoriaus dalyvavimas mokslinėse programose

Autorius dalyvavo:

- Vykdant Lietuvos valstybinio mokslo ir studijų fondo finansuotą projektą „Imčių teorijos naujų metodų kūrimas ir taikymas“ (2008-03-12–2009-08-31).
- 2007 m. rugpjūčio 13–17 d. Örebro mieste, Švedijoje vykuosiuose tęstinių duomenų analizės ir imčių koordinavimo klausimais kursuose doktorantams. Buvo įteikta pažyma už išklausytus kursus ir išlaikytus egzaminus temomis „Tęstinių duomenų analizė (*Analysis of Longitudinal Data*)“ ir „Imčių metodika (*Survey Methodology*)“.

Autorius stažavosi šioje užsienio mokslo įstaigoje:

- Stokholmo universitete (Švedija) nuo 2008 m. sausio 8 d. iki vasario 6 d. Stažuotės metu Statistikos katedros seminare perskaitytas pranešimas „Estimation of quantile in survey sampling“. Stažuotė finansuota iš projekto „Baltic-Nordic network co-operation on education and research in survey statistics“ lėšų.

Disertacijos struktūra

Disertaciją sudaro įvadas, trys pagrindiniai skyriai, išvados, naudotos literatūros sąrašas ir autoriaus publikacijų disertacijos tema sąrašas. Darbo apimtis yra 96 puslapiai, tekste panaudotos 176 formulės, 6 paveikslai, 10 lentelių. Rašant disertaciją buvo remtasi 56 literatūros šaltiniais.

1-asis skyrius yra apžvalginis. Šiame skyriuje apžvelgiami kitų autorių parašyti moksliniai darbai nagrinėjama tema.

2-ajame skyriuje pristatomi pirmieji autoriaus rezultatai: sudaryti baigtinės populiacijos tyrimo kintamojo reikšmių sudėtiniai santykiniai sumos įvertiniai, pateiktos apytikslių dispersijų išraiškos bet kokiam imties planui, o taip pat ir paparastosios atsitiktinės imties plano atveju. Taikant tokius įvertinius rotuojamai imčiai ir naudojant ankstesnių imties rinkimų duomenis, modeliuojant su realiais gyventojų užimtumo statistinio tyrimo duomenimis tikrinamas šių įvertinių efektyvumas. Modeliavimo rezultatai rodo, kad realiam gyventojų užimtumo statistinio tyrimo imties planui tikslinga naudoti papildomą informaciją santykiniam sumos įvertiniui. Nuo to įvertinių tikslumas tik pagerėja lyginant su tradiciniu imties planu pagrįstu baigtinės populiacijos tyrimo kintamojo sumos įvertiniu.

3-iojo skyriaus 1-ajame poskyryje sudaromas baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos imties planu pagrįstas įvertinys esant imties rotacijai. Taip pat sudaromi atitinkamai sudėtinis regresinis ir sudėtinis santykinis baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertinys. Pateikiamos sukonstruotų įvertinių apytikslės dispersijos išraiškos. Randami optimalūs baigtinės populiacijos tyrimo kintamojo pasiskirstymo funkcijos įvertiniai, kurių dispersija yra mažiausia. Atliekamas matematinis modeliavimas su realiais pajamų ir gyvenimo sąlygų statistinio tyrimo duomenimis, kurio metu sukonstruoti įvertiniai lyginami tarpusavyje.

3-iojo skyriaus 2-ajame poskyryje baigtinės populiacijos tyrimo kintamojo pasiskirstymo kvantilis vertinamas tiesiogiai, imant sudaryto sudėtinio santykinio pasiskirstymo funkcijos įvertinio, atvirkštinės funkcijos įvertinį. Remiantis matematiniu modeliavimu, sudarytas įvertinys lyginamas su standartiniu įvertiniu ir netiesioginiu santykiniu kvantilio įvertiniu, kuris gaunamas tiesiogiai iš žinomų rezultatų apie sumas ar vidurkius.

3-iojo skyriaus 3-iajame poskyryje nagrinėjami baigtinės populiacijos tyrimo kintamojo pasiskirstymo kvantilio pasikliautinojo intervalo sudarymo būdai, naudojant kartotinių imčių metodus. Sudaryti įvertiniai lyginami tarpusavyje remiantis matematiniu modeliavimu.

Padėka

Nuoširdžiai dėkoju darbo vadovams prof. habil. dr. Leonui Sauliui ir doc. dr. Danutei Krapavickaitei už įdomias idėjas ir naudingus patarimus, konsultacijas ir skirtą laiką; prof. Gunarui Kuldorfui (Gunnar Kulldorf, Umeå universitetas, Švedija) ir prof. Danieliui Torburnui (Daniel Thorburn, Stokholmo universitetas, Švedija), doc. dr. Danutei Krapavickaitei ir doc. dr. Aleksandrui Plikusui už stažuotes Švedijos universitetuose; kolegoms iš Vilniaus Gedimino technikos universiteto matematinės statistikos katedros už visokeriopą pagalbą ruošiant disertaciją.

Esu ypač dėkingas savo tėvams ir šeimos nariams už moralinį palaikymą ir paramą.

1

Įvairių baigtinės populiacijos parametrų įvertiniai

1.1. Žymenys ir pagrindinės sąvokos

Imčių tyrimai leidžia mums daryti išvadas apie baigtinės populiacijos charakteristikas, naudojant tik tos populiacijos poaibio duomenis. Tarkime, nagrinėjame iš N elementų sudarytą baigtinę populiaciją $\mathcal{U}$=$\{1, 2, \ldots, N\}$. Tegul y bus šioje populiacijoje apibrėžtas tyrimo kintamasis su nežinomomis reikšmėmis. Kintamojo y reikšmes žymėsime y_1, y_2, …, y_N ir laikysime pastoviomis. Turint baigtinę populiaciją, kurią sudaro N elementų, ir joje apibrėžtą kintamąjį $y : y_1, y_2, \ldots, y_N$, bet kurį baigtinės populiacijos parametrą θ galime užrašyti kaip kintamojo y reikšmių funkciją $\theta = \theta(y_1, y_2, \ldots, y_N)$. Pagrindinis uždavinys, nagrinėjant baigtines populiacijas, yra daryti išvadas apie įvairius tyrimo kintamojo y parametrus, tokius kaip kintamojo reikšmių suma ar vidurkis, taip pat ir sudėtingesnius parametrus – kintamojo skirstinio kvantilis, dispersija ir kt.

Vienas iš būdų tirti populiaciją ir daryti išvadas apie mus dominantį parametrą θ yra rinkti duomenis tik iš dalies populiacijos $\mathcal{U}$ elementų. Tą populiacijos dalį vadinsime *imtimi* ir žymėsime $\mathfrak{s} = \{1, 2, \ldots, n\}$. Imties elementų skaičius n vadinamas *imties dydžiu*. Imties elementų kintamojo y reikšmes žymėsime $y_1, y_2, \ldots, y_n$.

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 išrinkimo procedūrą, tenkinančią šiuos reikalavimus (Cochran 1977):

1. apibrėžiama aibė visų galimų skirtingų imčių $\mathfrak{S} = \{\mathfrak{s}_1, \mathfrak{s}_2, \dots \mathfrak{s}_v\}$;
2. kiekvienai aibės $\mathfrak{S}$ imčiai $\mathfrak{s}_i$, $i = 1, 2, \dots, v$, priskiriama žinoma imties išrinkimo tikimybė $p(\mathfrak{s}_i) > 0$, taip, kad $\sum_{k=1}^{v} p(\mathfrak{s}_k) = 1$;
3. kiekvienas populiacijos elementas priklauso bent vienai galimai imčiai;
4. imtis išrenkama taikant tokią atsitiktinę procedūrą, kai kiekviena aibės $\mathfrak{S}$ imtis $\mathfrak{s}_i$ išrenkama su nurodyta tikimybe $p(\mathfrak{s}_i)$, $i = 1, 2, \dots, v$.

Visos galimos imtys $\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_v$ ir tikimybės, su kuriomis jos gali būti išrinktos $p(\mathfrak{s}_1), p(\mathfrak{s}_2), \dots, p(\mathfrak{s}_v)$, apibrėžia tikimybinį skirstinį, vadinamą *imties planu*. Populiacijos elemento i priklausymo imčiai $\mathfrak{s}$ indikatorių žymėsime I_i, o jo reikšmes apibrėšime taip:

$$I_i = \begin{cases} 1, & \text{jei } i \in \mathfrak{s}; \\ 0, & \text{priešingu atveju.} \end{cases} \tag{1.1}$$

Populiacijos elementų priklausymo imčiai $\mathfrak{s}$ tikimybes galime apibrėžti taip:

- $\pi_i = \mathrm{P}(I_i = 1)$ pirmos eilės priklausymo imčiai tikimybė t.y. tikimybė, kad populiacijos elementas i priklauso imčiai
$$\pi_i = \sum_{\mathfrak{s}: i \in \mathfrak{s}} p(\mathfrak{s}),$$
visiems $i \in \mathcal{U}$.
- $\pi_{ij} = \mathrm{P}(I_i = 1, I_j = 1)$ antros eilės priklausymo imčiai tikimybė, t.y. tikimybė, kad du populiacijos elementai i ir j priklauso imčiai kartu
$$\pi_{ij} = \sum_{\mathfrak{s}: i,j \in \mathfrak{s}} p(\mathfrak{s}),$$
visiems $i, j \in \mathcal{U}$.

1.1 teiginys. *(Särndal* et al *1992) Nagrinėjamam imties planui $p(\mathfrak{s})$, indikatoriaus I_i* (1.1) *skaitinės charakteristikos yra:*

1. *vidurkis:* $\mathrm{E}(I_i) = \pi_i$,
2. *dispersija:* $\mathrm{D}(I_i) = \pi_i(1 - \pi_i)$,

3. *kovariacija:* $\mathrm{cov}(I_i,I_j) = \pi_{ij} - \pi_i\pi_j$, $i \neq j$, *visiems* $i,j \in \mathcal{U}$.

Išrinkus imtį ir surinkus imties elementų duomenis, populiacijos parametrų vertinimui pasitelkiamos tam tikros imties elementų funkcijos, kurios vadinamos parametrų įvertiniais. *Įvertinys* – tai taisyklė arba formulė, kuri nurodo, kaip turint imtį įvertinti popliacijos parametrą. Įvertinio reikšmė, apskaičiuota konkrečiai imčiai vadinama *įverčiu*. Parametro θ įvertinys žymimas $\hat{\theta} = \hat{\theta}(y_1, y_2 \dots y_n)$.

1.2. Įvertinių tikslumo matai

Turint kelis skirtingus populiacijos parametro θ įvertinius svarbu mokėti juos palyginti ir nustatyti, kuris iš jų bus „geresnis“. Įvertiniams palyginti naudojami įvairūs tikslumo matai:

- *poslinkis*, parodo kiek vidutiniškai parametro θ įvertinys $\hat{\theta}$ yra nutolęs nuo tikrosios parametro reikšmės
$$\mathrm{Posl}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta.$$
Kai $\mathrm{Posl}(\hat{\theta}) = 0$, tai įvertinys $\hat{\theta}$ vadinamas nepaslinktuoju parametro θ įvertiniu. Įvertinio nepaslinktumas yra svarbi jo savybė, tačiau ne visada galima sukonstruoti nepaslinktąjį įvertinį;
- *dispersija*, kuri apibūdina įvertinio $\hat{\theta}$ sklaidą apie jo vidutinę reikšmę
$$\mathrm{D}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \mathrm{E}(\hat{\theta}))^2,$$
tai paklaida, atsirandanti dėl imties atsitiktinumo;
- *vidutinė kvadratinė paklaida*, kuri apibūdina įvertinio $\hat{\theta}$ sklaidą apie tikrąją parametro θ reikšmę
$$\mathrm{VKP}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2 = \mathrm{D}(\hat{\theta}) + (\mathrm{Posl}(\hat{\theta}))^2.$$
Ši charakteristika vienu metu nusako du svarbiausius įvertinio $\hat{\theta}$ tikslumą apibūdinančius parametrus: dispersiją ir poslinkį. Kartais pasirenkamas toks parametro θ įvertinys $\hat{\theta}$, kuris turi nedidelį poslinkį, bet jo dispersija maža, nes tada jo vidutinė kvadratinė paklaida yra nedidelė;

- *santykinė standartinė paklaida* (variacijos koeficientas)

$$cv(\hat{\theta}) = \frac{\sqrt{\mathrm{D}(\hat{\theta})}}{\mathrm{E}(\hat{\theta})}.$$

Kuo šis koeficientas mažesnis, tuo įvertinys tikslesnis.

Yra ir kitų įvertinių tikslumo matų. Su keletu iš jų bus supažindinta sekančiuose skyreliuose. Kitame poskyryje bus nagrinėjamas universalus sumos įvertinys, tinkantis bet kokiam imties planui.

1.3. Horvico ir Tompsono sumos įvertinys

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{1, 2, \ldots, N\}$, sudarytą iš N elementų. Tegul y yra tyrimo kintamasis, įgyjantis reikšmes $y_1, y_2, \ldots, y_N$. Nagrinėsime baigtinės populiacijos tyrimo kintamojo reikšmių sumos

$$t_y = \sum_{i \in U} y_i,$$

vertinimą.

Horvitz, Thompson (1952) pasiūlė universalų nepaslinktąjį baigtinės populiacijos sumos įvertinį

$$\hat{t}_\pi = \sum_{i \in \mathfrak{s}} \frac{y_i}{\pi_i}, \qquad (1.2)$$

tinkantį vertinti baigtinės populiacijos tyrimo kintamojo reikšmių sumą bet kokio imties plano atžvilgiu. Kartais šis įvertinys vadinamas tyrimo kintamojo y π *sumos įvertiniu*, kadangi jame figūruoja populiacijos elemento i pirmos eilės priklausymo imčiai $\mathfrak{s}$ tikimybė.

1.2 teiginys. *(Horvitz, Thompson 1952) Tyrimo kintamojo y sumos įvertinio $\hat{t}_\pi$ savybės:*

1. *$\hat{t}_\pi$ yra nepaslinktasis sumos $t = \sum_{i \in \mathcal{U}} y_i$ įvertinys.*
2. *Šio įvertinio dispersija yra*

$$\mathrm{D}(\hat{t}_\pi) = \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}.$$

3. *Įvertinio $\hat{t}_\pi$ dispersijos įvertinys*

$$\widehat{\mathrm{D}}(\hat{t}_\pi) = \sum_{i \in \mathfrak{s}} \sum_{j \in \mathfrak{s}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}$$

yra nepaslinktasis, jei $\pi_{ij} > 0$ visiems $i,j \in \mathcal{U}$.

Dažnai literatūroje nagrinėjamas dar vienas baigtinės populiacijos parametras – populiacijos vidurkis μ_y. Išraišką (1.2) padauginus iš $\frac{1}{N}$ gaunamas baigtinės populiacijos tyrimo kintamojo y reikšmių nepaslinktasis vidurkio įvertinys $\hat{\mu}_\pi = \frac{1}{N}\hat{t}_\pi$ bet kokio imties plano atžvilgiu.

1.4. Sumos įvertiniai, naudojantys papildomą informaciją

Sudarant įvairių parametrų įvertinius dažnai įtraukiama žinoma papildoma informacija. Ji gali būti gaunama, pavyzdžiui iš ankstesnių tos pačios populiacijos tyrimų ar kitų duomenų bazių. Atliekant socialinius tyrimus galima remtis gyventojų surašymo ar demografiniais duomenims, atliekant įmonių tyrimus – statistinio įmonių registro duomenimis ir panašiai.

1.4.1. Santykinis įvertinys

Tarkime, kad nagrinėjamoje populiacijoje turime papildomą kintamąjį x, įgyjantį reikšmes $x_1, x_2, \ldots, x_N$ tiesiškai susijusį su tyrimo kintamuoju y, kurio visos populiacijos suma $t_x = \sum_{i=1}^N x_i$ yra žinoma. Gali būti tikslinga naudoti *santykinį įvertinį*

$$\hat{t}_y^{sant} = \frac{\hat{t}_{y\pi}}{\hat{t}_{x\pi}} t_x = \hat{r} t_x. \tag{1.3}$$

Čia $\hat{r} = \hat{t}_{y\pi}/\hat{t}_{x\pi}$, kur $\hat{t}_{y\pi}$ ir $\hat{t}_{x\pi}$ yra tyrimo kintamųjų, atitinkamai y ir x Horvico ir Tompsono nepaslinktieji sumų $t_y = \sum_{i=1}^N y_i$ ir $t_x = \sum_{i=1}^N x_i$ įvertiniai. Santykinis sumos įvertinys nėra nepaslinktasis. Kadangi įvertinys (1.3) yra netiesinė kelių sumų įvertinių funkcija, tai tikslios jo dispersijos išraiškos nėra. Apytikslei santykinio sumos įvertinio (1.3) dispersijos išraiškai rasti dažnai naudojamas Teiloro ištiesinimo (angl. *Taylor linearization*) metodas, kuris aprašytas 1.8.1 skyrelyje.

1.3 teiginys. *(Krapavickaitė, Plikusas 2005) Bet kokiam imties planui:*

1. *Populiacijos sumos santykinis įvertinys* $\hat{t}_y^{sant}$ (1.3) *gali būti apytiksliai užrašytas:*
$$\hat{t}_y^{sant} = \hat{r}t_x \approx rt_x + (\hat{t}_{y\pi} - r\hat{t}_{x\pi}).$$
2. *Šio įvertinio apytikslė dispersija yra*
$$\mathrm{AD}(\hat{t}_y^{sant}) = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}(\pi_{ij} - \pi_i\pi_j)\frac{y_i - rx_i}{\pi_i}\frac{y_j - rx_j}{\pi_j}.$$
3. *Įvertinio* $\hat{t}_y^{sant}$ *dispersijai vertinti patartina taikyti įvertinį*
$$\widehat{\mathrm{D}}(\hat{t}_y^{sant}) = \sum_{i\in\mathfrak{s}}\sum_{j\in\mathfrak{s}}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{y_i - \hat{r}x_i}{\pi_i}\frac{y_j - \hat{r}x_j}{\pi_j}.$$
Šis dipersijos įvertinys jau nėra nepaslinktasis.

Nors santykinis sumos įvertinys ir turi poslinkį, tačiau ne visada nepaslinktasis įvertinys turi mažiausią vidutinę kvadratinę paklaidą. Kartais geriau taikyti įvertinį turintį poslinkį, bet su mažesne dispersija. Santykinis sumos įvertinys sumažina sumos įvertinio vidutinę kvadratinę paklaidą, kai tyrimo kintamojo y reikšmės y_i yra apytiksliai proporcingos papildomo kintamojo x reikšmėms x_i, visiems $i \in \mathcal{U}$, t.y. tyrimo kintamojo y priklausomybė nuo x gali būti apytiksliai išreiškiama tiese, einančia per koordinačių pradžią.

1.4.2. Regresinis įvertinys

Galimos situacijos, kai baigtinės populiacijos tyrimo kintamasis y ir žinomas papildomas kintamasis x yra tiesiškai priklausomi, bet jų apytikslė priklausomybė išreiškiama tiese, nebūtinai einančia per koordinačių pradžią. Tokiose situacijose tikslinga naudoti *regresinį įvertinį*

$$\hat{t}_y^{reg} = \hat{t}_{y\pi} + \hat{b}(t_x - \hat{t}_{x\pi}), \tag{1.4}$$

Čia $\hat{t}_{y\pi}$ ir $\hat{t}_{x\pi}$ yra kintamųjų y ir x Horvico ir Tompsono nepaslinktieji sumų $t_y = \sum_{i=1}^N y_i$ ir $t_x = \sum_{i=1}^N x_i$ įvertiniai, o $\hat{b}$ regresijos koeficiento įvertinys, kuris parenkamas atsižvelgiant į imties planą.

1.4 teiginys. *(Krapavickaitė, Plikusas 2005) Bet kokiam imties planui*

1. *Populiacijos sumos regresinis įvertinys* $\hat{t}_y^{reg}$ (1.4) *nėra nepaslinktasis.*

2. *Šio įvertinio apytikslė dispersija yra*

$$\mathrm{AD}(\hat{t}_y^{reg}) = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}(\pi_{ij}-\pi_i\pi_j)\frac{y_i-a-bx_i}{\pi_i}\frac{y_j-a-bx_j}{\pi_j},$$

čia $a=\mu_y-b\mu_x$, $b=s_{xy}/s_x^2$.

3. *Įvertinio* $\hat{t}_y^{reg}$ *dispersijai vertinti rekomenduojamas įvertinys*

$$\widehat{\mathrm{D}}(\hat{t}_y^{reg}) = \sum_{i\in\mathfrak{s}}\sum_{j\in\mathfrak{s}}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\frac{y_i-\hat{a}-\hat{b}x_i}{\pi_i}\frac{y_j-\hat{a}-\hat{b}x_j}{\pi_j},$$

$$\hat{a}=\frac{1}{\widehat{N}}\Big(\sum_{i\in\mathfrak{s}}\frac{y_i}{\pi_i}-\hat{b}\sum_{i\in\mathfrak{s}}\frac{x_i}{\pi_i}\Big),\qquad \hat{b}=\frac{\sum_{i\in\mathfrak{s}}\frac{x_iy_i}{\pi_i}-\frac{1}{\widehat{N}}\hat{t}_{x\pi}\hat{t}_{y\pi}}{\sum_{i\in\mathfrak{s}}\frac{x_i^2}{\pi_i}-\frac{1}{\widehat{N}}\hat{t}_{x\pi}^2}.$$

Gali būti ir daugiau negu vienas papildomas kintamasis, susijęs su populiacijos tyrimo kintamuoju. Tarkime, kad turime ne vieną, o J toje pačioje populiacijoje apibrėžtų papildomų kintamųjų $x^{(1)}, x^{(2)}, \ldots, x^{(J)}$. Regresinius įvertinius, kuriuose naudojami keli papildomi kintamieji ir jų savybes nagrinėjo Cassel *et al* (1976), Isaki, Fuller (1982), Särndal *et al* (1989) ir kt. Pastaruoju metu praktikoje vis dažniau taikomi kalibruotieji sumų įvertiniai. Kalibravimo metodą baigtinės populiacijos tyrimo kintamojo sumos įvertiniui konstruoti pateikė Deville, Särndal (1992). Jie ištyrė tokio įvertinio savybes ir pateikė jo apytikslės dispersijos išraišką. Iš tikrųjų santykinis ir regresinis populiacijos sumų įvertiniai atitinkamai (1.3) ir (1.4) yra atskiri kalibruotųjų populiacijos sumų įvertinių atvejai.

1.5. Imties planai

Oficialiojoje statistikoje atliekant įvairius statistinius tyrimus iš baigtinės populiacijos renkamos imtys ir vertinami įvairūs parametrai naudojant įvairius imties planus. Šiame skyriuje pateiksime kelis imties planus, kurie bus naudojami disertacijoje įvairiems baigtinės populiacijos parametrams vertinti.

1.5.1. Paprastoji atsitiktinė imtis

Vienas iš paprasčiausių tikimybinių ėmimų yra *paprastoji atsitiktinė negrąžintinė imtis* (angl. *simple random sampling without replacement*). Tai tokia

n skirtingų elementų imtis iš N dydžio baigtinės populiacijos, kai bet kuris n skirtingų elementų rinkinys turi vienodą tikimybę būti išrinktas. Šiuo atveju kiekvienos n dydžio imties išrinkimo tikimybė yra $p(\mathfrak{s}) = \frac{1}{C_N^n}$. i-ojo populiacijos elemento ir dviejų populiacijos elementų i ir j priklausymo kartu kuriai nors imčiai tikimybės yra

$$\pi_i = \frac{n}{N}, \quad \pi_{ij} = \frac{n}{N}\frac{n-1}{N-1}, \quad i,j = 1,\dots,N, \quad i \neq j.$$

Tokią imtį žymėsime trumpiniu PAI.

1.5 teiginys. *(Särndal* et al *1992) Paprastosios atsitiktinės imties atveju*

1. *Horvico ir Tompsono sumos įvertinys $\hat{t}_\pi$ yra*

$$\hat{t}_{y,\mathrm{PAI}} = \frac{N}{n}\sum_{i\in\mathfrak{s}} y_i. \tag{1.5}$$

2. *Šio įvertinio dispersija yra*

$$\mathrm{D}(\hat{t}_{y,\mathrm{PAI}}) = N^2\left(1-\frac{n}{N}\right)\frac{s_y^2}{n}, \tag{1.6}$$

čia

$$s_y^2 = \frac{1}{N-1}\sum_{i\in U}(y_i-\mu)^2, \quad \mu = \frac{1}{N}\sum_{i\in U} y_i.$$

3. *Įvertinio $\hat{t}_{y,\mathrm{PAI}}$ dispersijos įvertinys*

$$\widehat{\mathrm{D}}(\hat{t}_{y,\mathrm{PAI}}) = N^2\left(1-\frac{n}{N}\right)\frac{\hat{s}_y^2}{n}, \tag{1.7}$$

čia

$$\hat{s}_y^2 = \frac{1}{n-1}\sum_{i\in\mathfrak{s}}(y_i-\overline{y})^2, \quad \overline{y} = \frac{1}{n}\sum_{i\in\mathfrak{s}} y_i,$$

yra nepaslinktasis.

Tikimybinis ėmimas, kai bet kuris n elementų rinkinys iš N dydžio populiacijos, kuris skiriasi nuo kitų elementų rinkinių ir kuriame gali būti pasikartojančių elementų, turi vienodą tikimybę būti išrinktas, vadinamas *paprastąja atsitiktine grąžintine imtimi*. Šiuo atveju kiekvienos n dydžio imties išrinkimo tikimybė yra $p(\mathfrak{s}) = \frac{1}{N^n}$.

1.5.2. Dviejų fazių ėmimas

Dviejų fazių ėmimas (angl. *two-phase sampling*) – tai toks tikimybinis ėmimas, kai iš išrinktos imties renkama kita, mažesnė imtis. Jos vadinamos atitinkamai pirmosios ir antrosios fazės imtimis. Dviejų fazių ėmimo schema nusakoma taip:

- pirmoje fazėje iš baigtinės populiacijos $\mathcal{U}$ pagal bet kokį imties planą išrenkama tikimybinė imtis $\mathfrak{s}^{(1)}$ su imties plano tikimybe $p(\mathfrak{s}^{(1)})$. Atitinkamai i-ojo ir j-ojo populiacijos elementų priklausymo šiai imčiai tikimybės žymimos

$$\pi_i^{(1)} = \mathrm{P}\big(\mathfrak{s}^{(1)} \subset \mathcal{U} : i \in \mathfrak{s}^{(1)}\big),$$
$$\pi_{ij}^{(1)} = \mathrm{P}\big(\mathfrak{s}^{(1)} \subset \mathcal{U} : i,j \in \mathfrak{s}^{(1)}\big),$$

  visiems $i,j \in \mathcal{U}$;

- antroje fazėje iš pirmosios fazės imties $\mathfrak{s}^{(1)}$ pagal bet kokį imties planą išrenkama tikimybinė imtis $\mathfrak{s}^{(2)}$ su imties plano tikimybe $p(\mathfrak{s}^{(2)}|\mathfrak{s}^{(1)})$. Atitinkamai i-ojo ir j-ojo populiacijos elementų priklausymo šiai imčiai sąlyginės tikimybės žymimos

$$\pi_{i|\mathfrak{s}^{(1)}}^{(2)} = \mathrm{P}\big(\mathfrak{s}^{(2)} \subset \mathfrak{s}^{(1)} : i \in \mathfrak{s}^{(2)}|\mathfrak{s}^{(1)}\big),$$
$$\pi_{ij|\mathfrak{s}^{(1)}}^{(2)} = \mathrm{P}\big(\mathfrak{s}^{(2)} \subset \mathfrak{s}^{(1)} : i,j \in \mathfrak{s}^{(2)}|\mathfrak{s}^{(1)}\big).$$

Dviejų fazių imties planui Horvico ir Tompsono tipo įvertinys populiacijos sumai vertinti nepatogus taikyti praktikoje, kadangi reikia žinoti antrosios fazės imties elementų sąlygines priklausymo imčiai tikimybes kiekvienos galimos pirmosios fazės imties realizacijos atveju. Galima alternatyva – naudoti π^* įvertinius (Särndal *et al* 1992), populiacijos sumai vertinti dviejų fazių imties atveju. Įvedama nauja tikimybė $\pi_i^* = \pi_i^{(1)}\pi_{i|\mathfrak{s}^{(1)}}^{(2)}$ ir sudaromas π^* sumos įvertinys.

1.6 teiginys. *Dviejų fazių imties atveju*

1. *Tyrimo kintamojo y sumos $t_y = \sum_{i\in\mathcal{U}} y_i$ sumos įvertinys*

$$\hat{t}_{\pi^*}^{(2)} = \sum_{i\in\mathfrak{s}} \frac{y_i}{\pi_i^{(1)}\pi_{i|s^{(1)}}^{(2)}} \tag{1.8}$$

yra nepaslinktasis.

2. *Šio įvertinio dispersija yra*

$$\begin{aligned}\mathrm{D}(\hat{t}^{(2)}_{\pi^*}) &= \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\big(\pi^{(1)}_{ij}-\pi^{(1)}_i\pi^{(1)}_j\big)\frac{y_i}{\pi^{(1)}_i}\frac{y_j}{\pi^{(1)}_j}\\ &+\mathrm{E}\sum_{i,j\in\mathfrak{s}^{(1)}}\big(\pi^{(2)}_{ij|\mathfrak{s}^{(1)}}-\pi^{(2)}_{i|\mathfrak{s}^{(1)}}\pi^{(2)}_{j|\mathfrak{s}^{(1)}}\big)\frac{y_i}{\pi^{(1)}_i\pi^{(2)}_{i|\mathfrak{s}^{(1)}}}\frac{y_j}{\pi^{(1)}_j\pi^{(2)}_{j|\mathfrak{s}^{(1)}}}.\end{aligned}$$

3. *Įvertinio $\hat{t}^{(2)}_{\pi^*}$ dispersijos įvertinys*

$$\begin{aligned}\widehat{\mathrm{D}}(\hat{t}^{(2)}_{\pi^*}) &= \sum_{i,j\in\mathfrak{s}^{(2)}}\frac{\pi^{(1)}_{ij}-\pi^{(1)}_i\pi^{(1)}_j}{\pi^{(1)}_{ij}\pi^{(2)}_{ij|\mathfrak{s}^{(1)}}}\frac{y_i}{\pi^{(1)}_i}\frac{y_j}{\pi^{(1)}_j}\\ &+\sum_{i,j\in\mathfrak{s}^{(2)}}\frac{\pi^{(2)}_{ij|\mathfrak{s}^{(1)}}-\pi^{(2)}_{i|\mathfrak{s}^{(1)}}\pi^{(2)}_{j|\mathfrak{s}^{(1)}}}{\pi^{(2)}_{ij|\mathfrak{s}^{(1)}}}\frac{y_i}{\pi^{(1)}_i\pi^{(2)}_{i|\mathfrak{s}^{(1)}}}\frac{y_j}{\pi^{(1)}_j\pi^{(2)}_{j|\mathfrak{s}^{(1)}}}\end{aligned}$$

yra nepaslinktasis.

1.1 Pastaba. Dviejų fazių ėmimo atveju sumos įvertinio (1.8) nepaslinktumas įrodomas ir jo dispersija užrašoma naudojant sąlyginių vidurkių ir dispersijų savybes:

$$\mathrm{E}(\hat{t}^{(2)}_{\pi^*}) = \mathrm{E}\big(\mathrm{E}(\hat{t}^{(2)}_{\pi^*}\mid\mathfrak{s}^{(1)})\big),$$

$$\mathrm{D}(\hat{t}^{(2)}_{\pi^*}) = \mathrm{D}\big(\mathrm{E}(\hat{t}^{(2)}_{\pi^*}\mid\mathfrak{s}^{(1)})\big)+\mathrm{E}\big(\mathrm{D}(\hat{t}^{(2)}_{\pi^*}\mid\mathfrak{s}^{(1)})\big).$$

Šios savybės gali būti naudojamos baigtinės populiacijos parametrų dispersijoms užrašyti naudojant ir kelių fazių ėmimą (Fuller 2003).

1.6. Imties rotacija

Nagrinėkime, baigtinę populiaciją $\mathcal{U} = 1, 2, \ldots, N$, sudarytą iš N elementų. Tegul y, įgyjantis reikšmes $y_1, y_2, \ldots, y_N$, yra tyrimo kintamasis. Nagrinėsime populiacijos parametrą $\theta = \theta(y_1, y_2, \ldots, N)$ priklausantį nuo tyrimo kintamojo y.

Daugelyje socialinių imčių, naudojant tą pačią populiaciją imtys renkamos pakartotinai ir to pačio tyrimo kintamojo reikšmės stebimos kiekvienoje iš jų. Tyrimai gali būti atliekami kas mėnesį, ketvirtį ar metus siekiant įvertinti tam tikrą baigtinės populiacijos parametrą θ. Kai tyrimo kintamojo reikšmės stebimos skirtingais laiko momentais gali būti sprendžiami šie uždaviniai:

1. baigtinės populiacijos parametro θ pokyčio vertinimas pereinant nuo vieno laiko momento prie kito;
2. baigtinės popliacijos parametro θ vertinimas apimant visus laiko momentus;
3. baigtinės populiacijos parametro θ vertinimas dabartiniu laiko momentu.

Daugelyje statistinių tyrimų domina baigtinės populiacijos parametro θ įvertis dabartiniu laiko momentu. Siekiant sumažinti įverčių svyravimus pereinant nuo vieno laiko momento prie kito keičiama ne visa imtis, o tiktai jos dalis, t.y. atliekama *imties rotacija*.

Nagrinėkime tokią N dydžio baigtinę populiaciją $\mathcal{U}$, kurios elementų skaičius ir patys elementai nekinta t laiko momentais, $t = 1, 2, \ldots$. Imties rotaciją galime nusakyti sekančiai:

- $(t-1)$ laiko momentu pagal tam tikrą imties planą iš baigtinės populiacijos $\mathcal{U}$ renkama n' dydžio imtis $\mathfrak{s}'$.
- t laiko momentu pagal tam tikrus imties planus m dydžio imtis $\mathfrak{s}_m$ renkama iš jau išrinktos n' dydžio imties $\mathfrak{s}'$ ir u dydžio imtis $\mathfrak{s}_u$ renkama iš $N - n'$ dydžio elementų aibės $\mathfrak{s}'^c = \mathcal{U} \setminus s^{'}$.

t laiko momentu turėsime išrinktą $n = m + u$ dydžio imtį $\mathfrak{s} = \mathfrak{s}_m \cup \mathfrak{s}_u$. Tokia ėmimo schema, kai baigtinės populiacijos parametras θ vertinamas naudojant t ir $t-1$ laiko momentų duomenis, vadinama *dviejų ėmimų schema* (angl. *two occasions scheme*).

Pirmasis dviejų ėmimų schemą nagrinėjo Jessen (1942). Kadangi imtis $\mathfrak{s}_m$ yra imties $\mathfrak{s}'$ poimtis, t.y. $\mathfrak{s}_m \subset \mathfrak{s}'$, todėl sudarant baigtinės populiacijos parametro θ įvertinius, galime vadovautis dviejų fazių ėmimo principais. Jessen (1942) baigtinės populiacijos tyrimo kintamojo y reikšmių vidurkiui $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$ vertinti, kai kiekviename iš ėmimų renkama paprastoji atsitikinė negražintinė imtis, pasiūlė sudėtinį įvertinį, imdamas tiesinę dviejų vidurkio įvertinių kombinaciją:

$$\overline{y}_t^* = \alpha \overline{y}_{m,t}^{reg} + (1-\alpha)\overline{y}_{u,t}, \tag{1.9}$$

$\overline{y}_{m,t}^{reg}$ – regresinis baigtinės populiacijos vidurkio įvertinys, sudaromas naudojant imčių $\mathfrak{s}'$ ir $\mathfrak{s}_m$ duomenis

$$\overline{y}_{m,t}^{reg} = \overline{y}_{m,t} + \hat{b}(\overline{y}_{n',t-1} - \overline{y}_{m,t-1}), \tag{1.10}$$

čia $\overline{y}_{m,t}$ ir $\overline{y}_{m,t-1}$ yra nepaslinktieji vidurkių $\mu_t = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}$ ir $\mu_{t-1} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}$ įvertiniai, $\hat{b}$ regresijos koeficiento įvertinys

$$\hat{b} = \frac{\sum_{i\in\mathfrak{s}_m}(y_{i,t}-\overline{y}_{m,t})(y_{i,t-1}-\overline{y}_{m,t-1})}{\sum_{i\in\mathfrak{s}_m}(y_{i,t-1}-\overline{y}_{m,t-1})^2}.$$

$\overline{y}_{u,t}$ – nepaslinktasis baigtinės populiacijos vidurkio įvertinys, sudaromas naudojant imties $\mathfrak{s}_u$ duomenis.

Jessen (1942) pasiūlė sudėtinio vidurkio įvertinio (1.9) dispersiją

$$D(\overline{y}_t^*) = \alpha^2 \mathrm{D}(\overline{y}_{m,t}^{reg}) + (1-\alpha)^2 \mathrm{D}(\overline{y}_{u,t}), \tag{1.11}$$

neįtraukdamas į jos išraišką kovariacijos nario $\mathrm{Cov}(\overline{y}_{m,t}^{reg}, \overline{y}_{u,t})$. Buvo surasta α konstanta, kuri minimizuoja dispersiją (1.11)

$$\alpha = \frac{\mathrm{D}(\overline{y}_{u,t})}{\mathrm{D}(\overline{y}_{u,t}) + \mathrm{D}(\overline{y}_{m,t}^{reg})}. \tag{1.12}$$

Patterson (1950), Eckler (1955) vertino populiacijos tyrimo kintamojo vidurkio pokytį pereinant nuo vieno laiko momento prie kito. Kulldorff (1963) nagrinėjo atskirą atvejį, jei dviejų ėmimų schemoje renkant imtis $\mathfrak{s}_m$ ir $\mathfrak{s}_u$ jų elementų apklausos kainos yra skirtingos. Sudėtinis vidurkio įvertinys (1.9), optimalaus imties dydžio parinkimas antrajame ėmime ieškant minimalios nagrinėjamo įvertinio dispersijos (1.11) buvo aptarti Cochran (1977), Sen (1971) kai kiekviename iš ėmimų renkama paprastoji atsitiktinė imtis. Sen *et al* (1975) pasiūlė naudojant imčių $\mathfrak{s}'$ ir $\mathfrak{s}_m$ duomenis, sudaryti santykinį vidurkio įvertinį. Jei koreliacijos koeficiento reikšmė $\rho(y_{t-1}, y_t)$ yra pakankamai didelė, tuomet sudėtinis santykinis vidurkio įvertinys turi mažesnę dispersiją negu nepaslinktasis tik imties planu pagrįstas vidurkio įvertinys.

Bendru atveju, kai kiekvienoje dviejų fazių ėmimo fazėje renkamos imtys pagal bet kokį imties planą, baigtinės populiacijos tyrimo kintamojo sumos sudėtinis regresinis įvertinys ir jo dispersija buvo aptarti Särndal *et al* (1992).

Caron, Ravalet (2000) apibrėžė nepaslinktąjį baigtinės populiacijos tyrimo kintamojo vidurkio $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$ įvertinį dviejų ėmimų atveju, kai kiekviename iš jų renkama paprastoji atsitiktinė imtis. Jų rezultatus galima suformuluoti sekančiu teiginiu.

1.7 teiginys. *Tarkime vidurkių įvertiniai* $\overline{y}_{m,\,t}$ *ir* $\overline{y}_{u,\,t}$ *yra nepaslinktieji,* $\overline{y}_{m,\,t} = \frac{1}{m}\sum_{i\in\mathfrak{s}_m} y_{i,\,t}$ *ir* $\overline{y}_{u,\,t} = \frac{1}{u}\sum_{i\in\mathfrak{s}_u} y_{i,\,t}$.

- *Vidurkio įvertinys* $\overline{y}_{n,\,t}$

$$\overline{y}_{n,\,t} = (1-k)\overline{y}_{m,\,t} + k\overline{y}_{u,\,t},$$

čia $k = \frac{m}{n}$, *yra nepaslinktasis;*

- *Įvertinio* $\overline{y}_{n,\,t}$ *dispersija yra*

$$\mathrm{D}(\overline{y}_{n,\,t}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n},$$

čia

$$s^2 = \frac{1}{n-1}\sum_{i\in\mathfrak{s}}(y_{i,\,t} - \overline{y}_{n,\,t})^2, \quad \overline{y}_{n,\,t} = \frac{1}{n}\sum_{i\in\mathfrak{s}} y_{i,\,t}.$$

Buvo įrodyta, kad jeigu kiekviename iš ėmimų renkama paprastoji atsitiktinė imtis, atitinkamai $\mathfrak{s}_m$ ir $\mathfrak{s}_u$, tuomet ir visa antrojo ėmimo imtis $\mathfrak{s}$ taip pat yra paprastoji atsitiktinė imtis išrinkta iš baigtinės populiacijos $\mathcal{U}$.

Imties rotacijos schemų gali būti įvairių, atskiruose statistiniuose tyrimuose taikomos skirtingos imties rotacijos schemos.

1.7. Baigtinės populiacijos sudėtingesnių parametrų vertinimas

Dažniausiai imčių teorijoje nagrinėjamas paprastų parametrų, tokių kaip suma ar vidurkis vertinimas baigtinėje populiacijoje. Sudėtingesni populiacijos parametrai, tokie kaip tyrimo kintamojo pasiskirstymo funkcija ar jo skirstinio kvantilis, taip pat yra sutinkami oficialiojoje statistikoje.

1.7.1. Imties planu pagrįstas pasiskirstymo funkcijos įvertinys

Tarkime, nagrinėjama baigtinė populiacija $\mathcal{U} = \{1, 2, \ldots, N\}$ su tyrimo kintamuoju y, kurio reikšmes žymėsime $y_1, y_2, \ldots, y_N$. Kiekvienai z reikšmei $(-\infty < z < \infty)$ baigtinės populiacijos tyrimo kintamojo y pasiskirstymo funkcija yra

$$F_y(z) = \frac{1}{N}\sum_{i\in U} I(y_i \le z), \tag{1.13}$$

su

$$I(y_i \le z) = \begin{cases} 1, & \text{jeigu } y_i \le z, \\ 0, & \text{jeigu } y_i > z. \end{cases}$$

Tarkime, kad tyrimo kintamojo y reikšmės y_i žinomos pagal tam tikrą imties planą išrinktos imties $\mathfrak{s}$ elementams $i \in \mathfrak{s}$. Atitinkamai i-ojo populiacijos elemento priklausymo šiai imčiai tikimybę pažymėkime π_i visiems $i \in \mathcal{U}$.

Vienas iš klasikinis pasiskirstymo funkcijos $F_y(z)$ įvertinių (Lahiri 1951) yra

$$\widehat{F}_y^{(1)}(z) = \frac{\sum_{i\in\mathfrak{s}} I(y_i \le z)/\pi_i}{\sum_{i\in\mathfrak{s}} 1/\pi_i}. \tag{1.14}$$

Chambers *et al* (1992) empiriškai parodė, kad su kai kuriomis pasirinktomis z reikšmėmis gaunamas didelis įvertinio (1.14) poslinkis.

Pasiskirstymo funkcijai $F_y(z)$ vertinti galima naudoti nepaslinktąjį Horvico ir Tompsono įvertinį

$$\widehat{F}_y(z) = N^{-1} \sum_{i\in\mathfrak{s}} I(y_i \le z)/\pi_i. \tag{1.15}$$

Kuk, Mak (1989) empiriškai palygino (1.14) ir (1.15) įvertinius. Jų eksperimento rezultatai parodė, kad Horvico ir Tompsono pasiskirstymo funkcijos $F_y(z)$ įvertinys (1.14) turi mažesnę vidutinę kvadratinę paklaidą negu pasiskirstymo funkcijos $F_y(z)$ įvertinys (1.15). Šie įvertiniai nenaudoja papildomų kintamųjų.

1.7.2. Papildomą informaciją naudojantys pasiskirstymo funkcijos įvertiniai

Tradicinis imties planu pagrįstas pasiskirstymo funkcijos $F_y(z)$ įvertinys (1.15) gali būti pagerintas taikant kitus įvertinius, naudojančius papildomus kintamuosius.

Tarkime, kad turime vieną papildomą kintamąjį x, kurio reikšmės yra žinomos visiems populiacijos elementams. Rao *et al* (1990) pasiūlė imties planu pagrįstus santykinį ir skirtuminį pasiskirstymo funkcijos $F_y(z)$ įvertinius.

Santykinis pasiskirstymo funkcijos $F_y(z)$ įvertinys apibrėžiamas taip:

$$\widehat{F}_y^{sant}(z) = \frac{1}{N} \frac{\sum_{i\in\mathfrak{s}} \pi_i^{-1} I(y_i \le z)}{\sum_{i\in\mathfrak{s}} \pi_i^{-1} I(\hat{r}x_i \le z)} \sum_{i\in U} \pi_i^{-1} I(\hat{r}x_i \le z), \tag{1.16}$$

čia $\hat{r} = (\sum_{i\in\mathfrak{s}} \pi_i^{-1} y_i)(\sum_{i\in\mathfrak{s}} \pi_i^{-1} x_i)$.

Tokio įvertinio vidutinė kvadratinė paklaida gali būti mažesnė negu $\widehat{F}_y(t)$ (1.15), kai tyrimo kintamojo y reikšmės yra atitinkamai proporcingos papildomo kintamojo x reikšmėms.

Skirtuminis pasiskirstymo funkcijos $F_y(z)$ įvertinys apibrėžiamas taip

$$\widehat{F}_y^{skirt}(z) = \frac{1}{N}\Bigg[\sum_{i\in\mathfrak{s}} \pi_i^{-1} I(y_i \le z) + \Bigg(\sum_{i\in U} \pi_i^{-1} I(\hat{R}x_i \le z) - \sum_{i\in\mathfrak{s}} \pi_i^{-1} I(\hat{R}x_i \le z)\Bigg)\Bigg]. \quad (1.17)$$

Abu pasiskirstymo funkcijos įvertiniai (1.16) ir (1.17) yra apytiksliai nepaslinktieji imties plano atžvilgiu. Didelių populiacijų ir didelių imčių atveju poslinkis yra nedidelis ir mažėja didinant imties dydį.

Chambers, Dunstan (1986), Wang, Dorfman (1996) nagrinėjo modeliu pagrįstus tyrimo kintamojo pasiskirstymo funkcijos įvertinius. Chambers *et al* (1992) lygino modeliu pagrįstus pasiskirstymo funckijos įvertinius su imties planu pagrįstais psiskirstymo funkcijos įvertiniais atsižvelgdami į jų pagrįstumą, poslinkį, dispersiją, tačiau nei vienam iš įvertinių pranašumo nesuteikė. Kuk (1993) nagrinėjo kitus pasiskirstymo funkcijos įvertinius, pagrįstus neparametrine regresija. Rao (1994), Rueda *et al* (2007) naudojo kalibravimo idėjas pasiskirstymo funkcijos įvertiniams sudaryti.

1.7.3. Imties planu pagrįsti kvantilių įvertiniai

Kvantilio vertinimas glaudžiai susijęs su pasiskirstymo funkcijos vertinimu. Kvantiliai naudojami ekonomikoje tyrinėjant ryšius tarp namų ūkio išlaidų maistui ir pajamų.

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{1,2,\ldots,N\}$ ir joje apibrėžtą tyrimo kintamąjį $y : y_1, y_2, \ldots, y_N$.

Baigtinėje populiacijoje apibrėžto kintamojo y skirstinio, q lygmens *kvantiliu* $q \in (0,1)$ vadiname

$$K_{yq} = \inf\{z : F_y(z) \ge q\}. \quad (1.18)$$

Atskiru atveju $q = 0{,}5$ lygmens kvantilis vadinamas *mediana*, kurią toliau žymėsime M_y, $M_y = \inf\{z : F_y(z) \ge 0{,}5\}$.

Norint įvertinti baigtinės populiacijos tyrimo kintamojo y skirstinio q lygio kvantilį pirmiausia reikia įvertinti pasiskirstymo funkciją $F_y(z)$, tuomet q lygmens kvantilis gali būti vertinamas taip:

$$\widehat{K}_{yq} = \inf\{z : \widehat{F}_y(z) \geq q\}. \tag{1.19}$$

Jei tyrimo kintamojo y pasiskirstymo funkcija $F_y(z)$ $(-\infty < z < \infty)$ yra tolydi ir didėjanti, tai jo skirstinio q, lygmens kvantilį K_{yq} galime apibrėžti taip

$$K_{yq} = F_y^{-1}(q), \tag{1.20}$$

čia F_y^{-1} yra funkcijos F_y atvirkštinė.

Tarkime pagal bet kokį imties planą renkama n dydžio imtis $\mathfrak{s}$. Imties elementų reikšmės išrikiuojamos didėjimo tvarka $y_1 \leq y_2 \leq \ldots \leq y_n$ ir kiekvienam iš elementų priskiriamos priklausymo imčiai $\mathfrak{s}$ tikimybės atitinkamai $\pi_1, \pi_2, \ldots, \pi_n$. Apibrėžiamos sukauptos sumos

$$B_k = \sum_{i=1}^{k} \frac{1}{\pi_i},$$

su kiekvienu $k = 1, \ldots, n$. Särndal *et al* (1992) tyrimo kintamojo skirstinio medianą M_y pasiūlė vertinti sekančiai:

$$\widehat{M}_y = \widehat{F}_y^{-1}(0.5) = \begin{cases} 0.5(y_{k-1} + y_k), & \text{jeigu } \frac{B_k}{\widehat{N}} = 0.5; \\ y_k, & \text{jeigu } \frac{B_{k-1}}{\widehat{N}} < 0.5 < \frac{B_k}{\widehat{N}}, \end{cases} \tag{1.21}$$

su kiekvienu $k \in \mathfrak{s}$, $\widehat{N} = \sum_{\mathfrak{s}} 1/\pi_k$. Tokiu būdu galime vertinti, bet kurį q lygio kvantilį K_{yq}.

Tokį kvantilio vertinimo būdą, kai pirma yra įvertinama pasiskirstymo funkcija, o po to vertinamas kvantilis imant jos atvirkštinę vadinkime tiesioginiu. Baigtinės populiacijos tyrimo kintamojos skirstinio medianą, tiesiogiai vertino Gross (1980) atsižvelgdamas tik į tyrimo kintamąjį, nenaudojant jokios papildomos informacijos. Silverman (1986), Lohr (1999) pasiūlė keletą tyrimo kintamojo pasiskirstymo funkcijos $F_y(z)$ glodinimo būdų.

1.7.4. Papildomą informaciją naudojantys kvantilio įvertiniai

Literatūros šaltinių apie tyrimo kintamojo skirstinio medianos ir kitų kvantilių vertinimą naudojant papildomus kintamuosius yra žymiai mažiau negu vidurkių ir sumų atveju. Kvantilių vertinime papildoma informacija gali būti įtraukiama dviem būdais:

1. konstruojant baigtinės populiacijos pasiskirstymo funkcijos įvertinius, kurie aprašyti 1.7.2 skyelyje. Tuomet kvantilis gali būti vertinamas tiesiogiai imant pasiskirstymo funkcijos atvirkštinę;
2. konstruojant netiesioginius kvantilių įvertinius taikant santykinius, skirtuminius ir regresinius įvertinius.

Tarkime, nagrinėjamoje populiacijoje turime papildomą kintamąjį x, įgyjantį reikšmes $x_1, x_2 \ldots, x_N$, kurio visos populiacijos mediana M_x yra žinoma.

Kuk, Mak (1989) apibrėžė baigtinės populiacijos tyrimo kintamojo medianos M_y santykinį įvertinį

$$\widehat{M}_y^{sant} = \frac{\widehat{M}_y}{\widehat{M}_x} M_x, \tag{1.22}$$

čia $\widehat{M}_x = \inf(z : \widehat{F}_x(z) \geq 0.5)$ ir $\widehat{M}_y = \inf(z : \widehat{F}_y(z) \geq 0.5)$, o $\widehat{F}_x(z)$, $\widehat{F}_y(z)$ yra atitinkamai papildomo kintamojo x ir tyrimo kintamojo y nepaslinktieji pasiskirstymo funkcijų įvertiniai.

Baigtinės populiacijos tyrimo kintamojo skirstinio q lygio kvantiliui K_{yq} vertinti gali būti naudojami įvairūs skirtuminiai įvertiniai. Pagrindinė idėja – naudojant tokius įvertinius yra išnaudoti turimą papildomą informaciją siekiant sumažinti papildomos informacijos nenaudojančio standartinio q lygio kvantilio įvertinio $\widehat{K}_{yq}$ (1.19) vidutinę kvadratinę paklaidą. Skirtuminių įvertinių klasė yra gaunama iš bendros išraiškos:

$$\widehat{K}_{yq}^{skirt} = \widehat{K}_{yq} + b(K_{xq} - \widehat{K}_{xq}), \tag{1.23}$$

čia b bet koks koeficientas. Atskiru atveju, kai $b = \widehat{K}_{yq}/\widehat{K}_{xq}$ gaunamas santykinis q lygio kvantilio įvertinys (Rao *et al* 1990).

Rao *et al* (1990) pasiūlė imties planu pagrįstą skirtuminį baigtinės populiacijos kvantilio įvertinį, koeficientą b pakeisdami kintamųjų y ir x vidurkių įvertinių santykiu

$$\widehat{K}_{yq}^{skirtR} = \widehat{K}_{yq} + \widehat{R}(K_{xq} - \widehat{K}_{xq}), \tag{1.24}$$

čia $\widehat{R} = \sum_{i\in s} y_i/\pi_i / \sum_{i\in s} x_i/\pi_i$. Šis kvantilio įvertinys sutampa su tikruoju kvantiliu K_{yq}, jei kintamasis y yra proporcingas kintamajam x. Šiuo atveju įvertinio (1.24) dispersija tampa lygi 0.

Rueda *et al* (2003) pasiūlė regresinį baigtinės populiacijos kvantilio įvertinį, koeficientą b pakeisdami tiesinės regresijos koeficento įvertiniu tarp tyrimo kintamojo y ir papildomo kintamojo x

$$\widehat{K}_{yq}^{reg} = \widehat{K}_{yq} + \hat{c}(K_{xq} - \widehat{K}_{xq}), \tag{1.25}$$

čia

$$\hat{c} = \frac{\sum_{i\in s} x_i y_i/\pi_i}{\sum_{i\in s} x_i^2/\pi_i}.$$

Kuk (1988) pasiūlė kvantilio įvertinį nelygių tikimybių imties planui. Branduolinius kvantilio įvertinius nagrinėjo Sheather, Marron (1990), Kuk (1993). Harms, Dushesne (2006), Rueda *et al* (2007) kvantilį vertino naudodami kalibravimo idėjas. Pagrindinis autorių tikslas buvo pasiūlyti kalibruotą įvertinį kvantiliui bet kokiam imties planui, kurį būtų patogu taikyti.

Rueda *et al* (2007) nagrinėjo dviejų ėmimų schemą, kuri aprašyta 1.6 skyrelyje, kai kiekviename iš ėmimų renkamos imtys pagal bet kokį imties planą. Tyrimo kintamojo y reikšmes pirmajame ėmime pažymėkime x_i, o jau pasikeitusias jo reikšmes antrajame ėmime žymėkime y_i. Buvo pasiūlyti netiesioginis sudėtinis santykinis ir netiesioginis sudėtinis regresinis baigtinės populiacijos tyrimo kintamojo pasiskirstymo q lygio kvantilio įvertiniai.

Sudėtinis santykinis q lygio kvantilio įvertinys buvo apibrėžtas taip:

$$\widehat{K}_{yq}^{sant} = \omega \widehat{K}_{m,yq}^{sant} + (1-\omega)\widehat{K}_{u,yq}, \tag{1.26}$$

čia ω, $\omega \in (0, 1)$ – laisvai pasirenkama konstanta,

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

atitinkamai $\widehat{K}_{m,yq} = \widehat{F}_{m,y}^{-1}(q)$, $\widehat{K}_{m,xq} = \widehat{F}_{m,x}^{-1}(q)$, $\widehat{K}_{n',xq} = \widehat{F}_{n',y}^{-1}(q)$, $\widehat{K}_{u,yq} = \widehat{F}_{u,y}^{-1}(q)$. Šis sudėtinis santykinis q lygio kvantilio įvertinys yra patogus taikyti praktikoje, kadangi nesunkiai apskaičiuojamas. Sudėtinis regresinis q lygio kvantilio įvertinys buvo apibrėžtas taip

$$\widehat{K}_{yq}^{reg} = \omega \widehat{K}_{m,yq}^{reg} + (1-\omega)\widehat{K}_{u,yq} \tag{1.27}$$

čia

$$\widehat{K}_{m,yq}^{reg} = \widehat{K}_{m,yq} + \hat{b}(\widehat{K}_{n',xq} - \widehat{K}_{m,xq}), \quad \hat{b} = \frac{\widehat{\mathrm{Cov}}(\widehat{K}_{m,yq}, \widehat{K}_{m,xq})}{\hat{\mathrm{D}}(\widehat{K}_{m,xq})}.$$

Straipsnio autoriai gauna optimalias ω reikšmių išraiškas. Tai reiškia, kad su tomis reikšmėmis įvertinių (1.26) ir (1.27) dispersija yra minimizuojama. Netiesioginius kvantilio įvertinius (1.26) ir (1.27) paprastosios atsitiktinės imties atveju užrašė ir nagrinėjo Martinez *et al* (2005). Montoya *et al* (2008) savo darbe konstruodami kvantilio įvertinius naudojo visrakčio metodą, kuris aptartas 1.8.2 skyrelyje.

1.8. Sudėtingesnių įvertinių dispersijų vertinimas

Ne visiems įvertiniams įmanoma surasti tikslią dispersijos išraišką. Tačiau panaudojus atitinkamus metodus įmanoma surasti apytiksles dispersijas. Šiame skyriuje glaustai aptarsime keletą metodų: *Teiloro ištiesinimo*, *visrakčio* (angl. *jackknife*), *savirankos* (angl. *bootstrap*), skirtų sudėtingesnių įvertinių dispersijoms vertinti.

1.8.1. Apytikslė įvertinio dispersija ir Teiloro ištiesinimo metodas jai rasti

Teiloro ištiesinimo metodas gali būti taikomas netiesinių sumų atžvilgiu populiacijos parametrų apytikslėms dispersijoms apskaičiuoti.

Tarkime, kad baigtinėje populiacijoje $\mathcal{U}$ turime p apibrėžtų tyrimo kintamųjų $y^{(1)}, \ldots, y^{(p)}$ ir populiacijos parametrą θ, kuris yra šių kintamųjų funkcija:

$$\theta = \theta(y^{(1)}, \ldots, y^{(p)}).$$

Nagrinėkime baigtinės populiacijos parametro θ įvertinį, kuris gali būti išreikštas sumų $t_k = \sum_{i=1}^{N} y_i^{(k)}$, $k = 1, \ldots p$, nepaslinktųjų įvertinių $\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi}$ funkcija

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

Laikysime, kad įvertiniai $\hat{t}_{k\pi}$ yra pagrįsti imties planu, turintys pavidalą

$$\hat{t}_{k\pi} = \sum_{i \in \mathfrak{s}} \frac{y_i^{(k)}}{\pi_i}.$$

Jei dominantis parametro θ įvertinys yra netiesinė p populiacijos sumų įvertinių funkcija, tai įvertinio $\hat{\theta} = f(\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi})$ dispersiją tiksliai užrašyti dažniausiai yra sunku arba iš vis neįmanoma. Tarkime, kad parametro θ įvertinys $\hat{\theta} = f(\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi})$ yra tolydi funkcija taško $(t_1, \ldots, t_p)$ aplinkoje ir tame taške turinti dalines išvestines. Tokiam įvertiniui patogu taikyti Teiloro formulę taške $(t_1, \ldots, t_p)$, imant tik pirmosios eilės narius:

$$\hat{\theta} = f(\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi}) \approx \hat{\theta}_0 = a_0 + \sum_{k=1}^{p} a_k(\hat{t}_{k\pi} - t_k), \tag{1.28}$$

čia

$$a_0 = f(t_1, \ldots, t_p), \quad a_k = \frac{\partial f(\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi})}{\partial \hat{t}_{k\pi}}\bigg|_{(\hat{t}_{1\pi}, \ldots, \hat{t}_{p\pi}) = (t_1, \ldots, t_p)}, k = 1, \ldots p.$$

Įvertinio $\hat{\theta}$ tiesinės dalies $\hat{\theta}_0$ dispersija $\mathrm{D}(\hat{\theta}_0)$ yra vadinama įvertinio $\hat{\theta}$ *apytiksle dispersija* (Särndal *et al* 1992) ir žymima $\mathrm{AD}(\hat{\theta})$:

$$\mathrm{AD}(\hat{\theta}) = \mathrm{D}\left(\sum_{k=1}^{p} a_k \hat{t}_{k\pi}\right) = \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{ij} - \pi_i \pi_j) \frac{u_i}{\pi_i} \frac{u_j}{\pi_j}, \tag{1.29}$$

čia $u_i = \sum_{k=1}^{p} a_k y_{ki}$.

Dispersijos $\mathrm{D}(\hat{\theta})$ įvertiniu imsime

$$\widehat{\mathrm{D}}(\hat{\theta}) = \sum_{i \in \mathfrak{s}} \sum_{j \in \mathfrak{s}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{\hat{u}_i}{\pi_i} \frac{\hat{u}_j}{\pi_j}, \tag{1.30}$$

gaunamą vertinant $\mathrm{AD}(\hat{\theta})$ nepaslinktuoju įvertiniu. Čia $\hat{u}_i = \sum_{k=1}^{p} \hat{a}_k y_{ki}$ ir $\hat{a}_k$ yra a_k koeficiento įvertinys.

Teiloro ištiesinimo metodas tinka visiems imčių planams ir įvairių parametrų įvertiniams, bet jį taikant reikia ieškoti dalinių išvestinių sumų įvertinių atžvilgiu.

1.8.2. Visrakčio metodas

Visrakčio metodą kaip priemonę įvertinio poslinkiui sumažinti pasiūlė Quenoille (1949). Tukey (1958) pastebėjo, kad šis metodas gali būti naudojamas įvertinių dispersijoms vertinti. Durbin (1959) buvo pirmasis autorius panaudojęs šią teoriją baigtinėje populiacijoje.

Tarkime iš baigtinės populiacijos $\mathcal{U}$ renkama n dydžio imtis $\mathfrak{s}$ pagal tam tikrą imties planą, su žinoma imties išrinkimo tikimybe $p(\mathfrak{s})$.

Pažymėkime $\hat{\theta}^*$ vertinamo parametro θ įvertinį. Taikant visrakčio metodą iš baigtinės populiacijos $\mathcal{U}$ išrinkta fiksuota, tarkime, n dydžio imtis s skaidoma į K vienodo dydžio atsitiktinių grupių. Kiekvienai grupei $k = 1, \dots, K$ apibrėžiamas parametro θ įvertinys $\hat{\theta}^{(k)}$, kurio pavidalas toks pat kaip ir $\hat{\theta}^{*}$, tik įvertinys $\hat{\theta}^{(k)}$ skaičiuojamas naudojant imties $\mathfrak{s}$ duomenis, be elemento k. Apibrėžiamos *pseudoreikšmės*:

$$\hat{\theta}^{k} = K\hat{\theta}^{*} - (K-1)\hat{\theta}^{(k)},$$

tuomet parametro θ visrakčio įvertinys yra

$$\hat{\theta}^{vis} = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}^{k}. \tag{1.31}$$

Pastarojo įvertinio dispersijos įvertiniai

$$\widehat{\mathrm{D}}_{vis1} = \frac{1}{n(n-1)}\sum_{a=1}^{n}(\hat{\theta}^{(a)} - \hat{\theta}^{vis})^2, \tag{1.32}$$

$$\widehat{\mathrm{D}}_{vis2} = \frac{1}{n(n-1)}\sum_{a=1}^{n}(\hat{\theta}^{(a)} - \hat{\theta}^{*})^2, \tag{1.33}$$

yra vadinami visrakčio dispersijos įvertiniais. Praktikoje taikant abi šias formules gaunami panašūs rezultatai.

Daugelis straipsnių, kuriuose naudojamas visrakčio metodas skirti baigtinės populiacijos parametrų, tokių kaip vidurkis ar suma, vertinimui (Shao, Tu 1995; Shao, Wu 1989). Wu, Sitter (2001) taikė visrakčio metodą įvairių pasiskirstymo funkcijos įvertinių dispersijoms vertinti. Naudodami šį metodą galime vertinti ir sudėtingų populiacijos parametrų įvertinių dispersijas.

1.8.3. Savirankos metodas

Efron (1979) pasiūlė savirankos metodą įvairių baigtinės populiacijos parametrų savybėms tirti. Šio metodo idėja – iš pradinės imties pakartotinai renkant poimčius sukurti tokias duomenų aibes, iš kurių būtų galima įvertinti populiacijos parametrų įvertinių sklaidos charakteristikas. Taikant šį metodą iš pradinės imties daug kartų išrenkama nauja imtis ir parametro vertinimo procedūra kartojama.

Tarkime, nagrinėjama baigtinė populiacija $\mathcal{U}$, kurios elementų tyrimo kintamojo y reikšmės yra $y_1, \dots, y_N$. Tarkime iš baigtinės populiacijos $\mathcal{U}$ renkama n dydžio imtis $\mathfrak{s}$ pagal tam tikrą imties planą, su žinoma imties išrinkimo tikimybe $p(\mathfrak{s})$. Pažymėkime $\hat{\theta}^*$ vertinamo parametro θ įvertinį. Savirankos metodo taikymas gali būti nusakytas tokia žingsnių seka:

Iš turimos n dydžio imties $\mathfrak{s}$ renkamas m dydžio poimtis, kuris vadinamas *savirankos imtimi*. Kartodami tokių poimčių rinkimą B kartų, gausime populiacijos parametro θ įvertinius $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$.

Tuomet paramtero θ įvertinio $\hat{\theta}$ poslinkis ir dispersija atitinkamai įvertinami:

$$\widehat{\mathrm{Posl}}_*(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}^{*b} - \overline{\hat{\theta}}^{*b}), \quad \widehat{\mathrm{D}}_*(\hat{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{*b} - \overline{\hat{\theta}}^{*b})^2,$$

čia

$$\overline{\hat{\theta}}^{*b} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}.$$

Kai imties dydis nedidelis, tradicinį savirankos metodą reikėtų modifikuoti siekiant išvengti juo gaunamo dispersijos įvertinio poslinkio. Rao, Wu (1988) pasiūlė *mastelio pakeitimo savirankos* (angl. *rescaling bootstrap*) metodą. Pagrindinė šio metodo idėja tyrimo kintamojo y reikšmes y_i patekusiais į savirankos imtį, transformuoti pakeičiant mastelį

$$\widetilde{y}_i = \overline{y} + \sqrt{\left(1 - \frac{n}{N}\right)}\frac{\sqrt{m}}{\sqrt{n-1}}\Big(y_i - \overline{y}\Big), \quad i = 1, \ldots, m.$$

Stein (1989) nagrinėjo savirankos metodo pritaikymo galimybes įvairių parametrų įvertinių dispersijoms vertinti. Shao (2003) apžvelgia savirankos metodo modifikacijų vystymąsi. Taikant šį metodą reikia daug kompiuterinio laiko.

1.9. Kvantilio pasikliautinasis intervalas ir jo įvertiniai

Praktikoje dažnai naudojamas baigtinės populiacijos parametro θ įvertinio $\hat{\theta}$ paklaidos matas – *pasikliautinasis intervalas*. Tai toks atsitiktinis intervalas $(\hat{\theta}_A, \hat{\theta}_V)$, kuriame su tikimybe $1-\alpha$, $\alpha > 0$ yra tikroji parametro θ reikšmė:

$$\mathrm{P}(\theta \in (\hat{\theta}_A, \hat{\theta}_V)) = 1 - \alpha, \tag{1.34}$$

čia skaičius $1-\alpha$ vadinamas *pasikliovimo lygmeniu*, o skaičius α vadinamas *klaidos tikimybe*. Galimas simetrinis pasikliautinasis intervalas $(\hat{\theta}-d,\hat{\theta}+d)$, kai absoliučioji paklaida $d>0$, su tikimybe $1-\alpha$ dengiantis tikrąją parametro θ reikšmę, kuris apibrėžiamas taip

$$\mathrm{P}(|\,\hat{\theta}-\theta\,|<d)=1-\alpha,$$

arba, kai $\mathrm{D}(\hat{\theta})>0$,

$$\mathrm{P}(|\,\hat{\theta}-\theta\,|<d)=\mathrm{P}\left(\frac{|\,\hat{\theta}-\theta\,|}{\sqrt{\mathrm{D}(\hat{\theta})}}<\frac{d}{\sqrt{\mathrm{D}(\hat{\theta})}}\right)=1-\alpha. \tag{1.35}$$

Dažnai tokie pasikliautinieji intervalai konstruojami darant prielaidą, kad įvertinio $\hat{\theta}$ skirstinys yra normalusis. Tada pasikliautinasis intervalas įgyja pavidalą:

$$(\hat{\theta}-d,\hat{\theta}+d)=\left(\hat{\theta}-z_{\alpha/2}\sqrt{\mathrm{D}(\hat{\theta})},\hat{\theta}+z_{\alpha/2}\sqrt{\mathrm{D}(\hat{\theta})}\right), \tag{1.36}$$

čia $z_{\alpha/2}$ yra standartinio normaliojo skirstinio $1-\alpha/2$ lygmens kvantilis. Kadangi parametro įvertinio $\hat{\theta}$ dispersija $\mathrm{D}(\hat{\theta})$ paprastai nežinoma, ji keičiama šios dispersijos įvertiniu ir gaunamas pasikliautinojo intervalo įvertinys:

$$(\hat{\theta}-d,\hat{\theta}+d)=\left(\hat{\theta}-z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}(\hat{\theta})},\hat{\theta}+z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}(\hat{\theta})}\right).$$

Nagrinėjant baigtines populiacijas įvertinio $\hat{\theta}$ tikimybinis skirstinys yra diskretus, todėl jis negali būti normalusis. Didelėms populiacijoms ir dideliems imties dydžiams daugelis taikomų įvertinių turi apytiksliai normalųjį skirstinį. Hajek (1960) nagrinėdamas baigtinės populiacijos vidurkio μ įvertinį $\hat{\mu}=\overline{y}$ paprastosios atsitiktinės negrąžintinės imties atveju įrodė, kad

$$\frac{\overline{y}-\mu}{\sqrt{\left(1-\frac{n}{N}\right)\frac{\hat{s}^2}{n}}}$$

skirstinys artėja į standartinį normalųjį, kai populiacijos dydis N ir imties dydis n tam tikru būdu artėja į begalybę.

Daugelyje darbų tokios kintamojo padėties charakteristikos, kaip mediana, kvartilis, decilis ir kiti nagrinėjamos su prielaida, kad jų įvertinio skirstinys yra normalusis. Woodurff (1952) apibrėžė normaliuoju skirstiniu pagrįstą baigtinės populiacijos tyrimo kintamojo medianos $M_y = K_{y0,5}$ pasikliautinojo intervalo įvertinį taip:

$$\big(\widehat{M}_y^A, \widehat{M}_y^V\big), \tag{1.37}$$

čia

$$\widehat{M}_y^A = \inf\{z : \widehat{F}(z) \geq 0{,}5 - z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}\widehat{F}_y(\widehat{M}_y)}\},$$
$$\widehat{M}_y^V = \inf\{z : \widehat{F}(z) \geq 0{,}5 + z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}\widehat{F}_y(\widehat{M}_y)}\},$$

ir

$$\mathrm{P}(\widehat{M}_y^A \leq M_y \leq \widehat{M}_y^V) = 1 - \alpha.$$

Nomaliuoju skirstiniu pagrįstas baigtinės populiacijos tyrimo kintamojo skirstinio q lygmens kvantilio K_{yq} $(1-\alpha)$ lygmens pasikliautinojo intervalo įvertinys apibrėžiamas (Sitter, Wu 2001) taip:

$$\big(\widehat{K}_{yq}^A, \widehat{K}_{yq}^V\big) = \left(\widehat{F}_y^{-1}\big(q - z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}\widehat{F}_y(\widehat{K}_{yq})}\big), \big(\widehat{F}_y^{-1}\big(q + z_{\alpha/2}\sqrt{\widehat{\mathrm{D}}\widehat{F}_y(\widehat{K}_{yq})}\big)\right). \tag{1.38}$$

Tokio tipo pasikliautinojo intervalo įvertinys turi 2 ypatybes:

1. Kai kuriais atvejais nėra įmanoma tiesiogiai įvertinti tyrimo kintamojo pasikirstymo funkcijos dispersijos.
2. Pasiskirstymo funkcijos $F(z)$ įvertinio $\widehat{F}(z)$ apytikslio normališkumo prielaida gali būti netenkinama daugelyje praktinių situacijų, kai imties dydžiai yra maži.

Francisko, Fuller (1959) nagrinėdamas tyrimo kintamojo pasiskirstymo funkcijos $F_y(z)$ įvertinį $\widehat{F}_y(z)$ įrodė, kad

$$\frac{\widehat{F}_y(z) - F_y(z)}{\sqrt{\widehat{\mathrm{D}}(\widehat{F}_y(z))}},$$

skirstinys artėja į standartinį normalųjį, kai populiacijos dydis ir imties dydis tam tikru būdu artėja į begalybę.

Kaip alternatyva normaliuoju skirstiniu pagrįstam kvantilio pasikliautinajam intervalui gali būti konstruojami intervalai naudojant imčių perrinkimo metodus dispersijai vertinti. McCarthy (1993) nagrinėjo keletą pasikliautinojo intervalo sudarymo procedūrų baigtinės populiacijos tyrimo kintamojo skirstinio medianai, naudojant paprastąją atsitiktinę negrąžintinę imtį. DiCiccio, Efron (1996) parodė, kad panaudojus savirankos metodą gali būti gaunami pakankamai tikslūs pasikliautinieji intervalai. Tačiau šiuo metodu skaičiavimas trunka labai ilgai.

Vieni pirmųjų darbų Šiaurės-Baltijos šalyse, kuriuose nagrinėjamas kvantilis buvo atlikti Švedijos statistikės Orusild (1998), Orusild (1999). Ji nagrinėjo parametrus, kurie gali būti išreikšti kvantilių funkcijomis ir pasiūlė pasikliautinojo intervalo sudarymo procedūrą pagrindinėms kvantilių funkcijoms.

1.10. Pirmojo skyriaus apibendrinimas. Uždavinių formulavimas

Apibendrinus šio skyriaus medžiagą, formuluojame šiuos uždavinius:

1. Sudaryti sudėtinius santykinius baigtinės populiacijos sumos įvertinius esant imties rotacijai. Užrašyti tokių įvertinių dispersijų išraiškas.
2. Sudaryti sudėtingesnių baigtinės populiacijos parametrų: tyrimo kintamojo pasiskirstymo funkcijos, kvantilio įvertinius esant imties rotacijai.
3. Sudaryti pasikliautinojo intervalo įvertinį baigtinės populiacijos kvantiliui naudojant imčių perrinkimo procedūras.
4. Modeliuojant palyginti sukonstruotus įvertinius tarpusavyje ir su standartiniais atitinkamų parametrų įvertiniais.

2

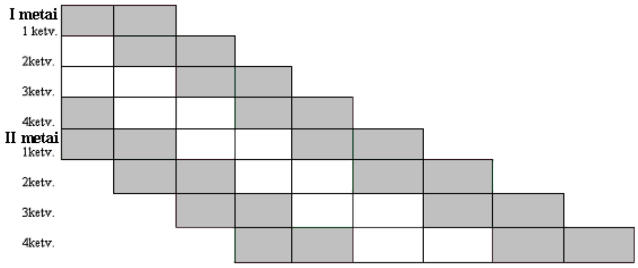
Baigtinės populiacijos sumos vertinimas esant imties rotacijai

Gyventojų užimtumo statistinis tyrimas šiuo metu vienas svarbiausių daugelio šalių oficialiosios statistikos socialinių tyrimų. Jis atliekamas ir Statistikos departamente prie Lietuvos respublikos vyriausybės. Jame pagal pasirinktą imties planą iš gyventojų registro išrinkus 15 m. ir vyresnių gyventojų imtį, iš jų surenkami duomenys ir vertinamas dirbančių gyventojų, bedarbių ir neaktyvių gyventojų skaičius Lietuvoje ir jos srityse. Gyventojų užimtumo statistinis tyrimas Lietuvoje atliekamas kiekvieną metų ketvirtį.

Kadangi norima sekti užimtumo rodiklių kitimą laike, tai, siekiant sumažinti įverčių svyravimą dėl imties pakeitimo, pereinant nuo vieno ketvirčio prie kito keičiama ne visa imtis, o tiktai jos ketvirtadalis, t. y. daroma imties rotacija. Šiek tiek supaprastinus situaciją, galima pasakyti, kad kiekvienas į imtį išrinktas asmuo įtraukiamas į dviejų iš eilės atliekamų tyrimų imtis, po to į dviejų tyrimų imtį jis neįtraukiamas ir vėl įtraukiamas į dviejų iš eilės atliekamų statistinių tyrimų imtis. Imties rotacijos schema vaizduojama 2.1 paveiksle.

Dėl imties rotacijos apie kiekvieną imties elementą, kuris įtrauktas į imtį jau nebe pirmą kartą, turimi ne tiktai tiriamo metų ketvirčio duomenys, bet ir anksčiau atlikto tyrimo duomenys.

Kol kas skaičiuojant užimtumo parametrų įverčius Statistikos departamente turima ankstesnių tyrimų informacija apie imties elementus nebuvo naudojama. Šiame skyriuje konstruosime sudėtinius santykinius baigtinės populiacijos



2.1 pav. Gyventojų užimtumo statistinio tyrimo imties rotacijos schema
Fig. 2.1. Sampling rotation scheme

sumos įvertinius naudodami kelių fazių ėmimą. Sukonstruotus įvertinius palyginsime su nepaslinktuoju Horvico ir Tompsono sumos įvertiniu empiriškai, atlikdami matematinį modeliavimą su realiais duomenimis.

2.1. Sumos vertinimas naudojant dviejų ėmimų schemą

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{1, \ldots, N\}$, iš kurios išrinkta tikimybinė imtis $\mathfrak{s}$ pagal imties planą $p(\mathfrak{s})$. Tarkime, kad y yra populiacijoje $\mathcal{U}$ apibrėžtas tyrimo kintamasis, kurio reikšmės $\{y_1, y_2, \ldots y_N\}$.

Sumai t_y vertinti naudosime dviejų iš eilės atliekamų statistinių tyrimų imtis, taikydami dviejų ėmimų schemą ir naudodami papildomą informaciją apie tyrimo kintamąjį y, žinomą iš ankstesnio tyrimo imties duomenų.

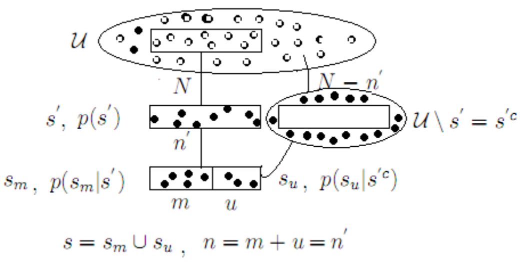
Tarkime, kad nagrinėjamos populiacijos $\mathcal{U}$ elementų skaičius ir patys elementai nekinta laike. Pirmuoju laiko momentu tyrimo kintamojo y reikšmes žymėsime x_i, o antruoju laiko momentu y_i.

Pirmasis ėmimas. Pirmuoju laiko momentu iš baigtinės populiacijos $\mathcal{U}$ pagal tam tikrą imties planą renkama $n^{'}$ dydžio imtis $\mathfrak{s}'$ su imties plano tikimybe $p(\mathfrak{s}')$. Atitinkamai i-ojo ir j-ojo populiacijos elemento priklausymo šiai imčiai tikimybes žymėsime $\pi_i' = \mathrm{P}(\mathfrak{s}' \subset \mathcal{U} : i \in \mathfrak{s}^{'})$, $\pi_{ij}^{'} = \mathrm{P}(\mathfrak{s}^{'} \subset \mathcal{U} : i \in \mathfrak{s}', j \in \mathfrak{s}')$, visiems $i,j \in \mathcal{U}$.

Antrasis ėmimas. Antruoju laiko momentu iš jau išrinktos $n^{'}$ dydžio imties $\mathfrak{s}'$ pagal tam tikrą imties planą renkama m dydžio imtis $\mathfrak{s}_m$, $\mathfrak{s}_m \subset \mathfrak{s}'$ su imties plano sąlygine tikimybe $p(\mathfrak{s}_m|\mathfrak{s}')$. Atitinkamas i-ojo ir j-ojo populiacijos elemento priklausymo imčiai $\mathfrak{s}_m$ sąlygines tikimybes žymėsime $\pi_{i|\mathfrak{s}'} = \mathrm{P}(\mathfrak{s}_m \subset \mathfrak{s}' : i \in \mathfrak{s}_m \mid \mathfrak{s}')$, $\pi_{ij|\mathfrak{s}'} = \mathrm{P}(\mathfrak{s}_m \subset \mathfrak{s}' : i \in \mathfrak{s}_m, j \in \mathfrak{s}_m \mid \mathfrak{s}')$.

Be to, iš $N - n'$ dyžio aibės $\mathfrak{s}'^c = \mathcal{U} \setminus \mathfrak{s}'$ elementų pagal tam tikrą imties planą renkama u dydžio imtis $\mathfrak{s}_u$. Imties $\mathfrak{s}_u$ plano sąlyginę tikimybę žymėsime $p(\mathfrak{s}_u|\mathfrak{s}'^c)$. Atitinkamas i-ojo ir j-ojo populiacijos elemento priklausymo šiai imčiai sąlygines tikimybes žymėsime $\pi_{i|\mathfrak{s}'^c} = \mathrm{P}(\mathfrak{s}_u \subset \mathfrak{s}'^c : i \in \mathfrak{s}_u \mid \mathfrak{s}'^c)$, $\pi_{ij|\mathfrak{s}'^c} = \mathrm{P}(\mathfrak{s}_u \subset \mathfrak{s}'^c : i \in \mathfrak{s}_u, j \in \mathfrak{s}_u \mid \mathfrak{s}'^c)$.

Tokiu būdu nagrinėjama $n = m+u$ elementų imtis $\mathfrak{s}_n$, kurią sudaro dviejų imčių sąjunga $\mathfrak{s}_m \cup \mathfrak{s}_u$. Imties $\mathfrak{s}_n$ išrinkimo procedūra pateikta 2.2 paveiksle.



2.2 pav. Dviejų ėmimų schema
Fig. 2.2. Two occasions scheme

Norėdami įvertinti užimtųjų gyventojų skaičių Lietuvoje, laikykime tyrimo kintamąjį y dvireikšmiu:

$$y_i = \begin{cases} 1, & \text{jei gyventojas } i \text{ užimtas,} \\ 0 & \text{priešingu atveju,} \end{cases}$$

$$x_i = \begin{cases} 1, & \text{jei gyventojas } i \text{ buvo užimtas,} \\ 0 & \text{priešingu atveju,} \end{cases}$$

$i \in U$. Tada užimtųjų gyventojų skaičius lygus kintamojo y reikšmių populiacijos sumai:

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

Vertinsime šią sumą naudodami nurodytos imties kintamojo y reikšmes y_i, $i \in \mathfrak{s}$, ir tyrimo kintamojo ankstesnio tyrimo reikšmes x_i, $i \in \mathfrak{s}'$.

Tiek imtis $\mathfrak{s}_m$, tiek ir $\mathfrak{s}_u$ yra dviejų fazių imtys. Toliau nagrinėsime kiekvieną iš imčių atskirai ir naudodami imčių duomenis sudarysime tyrimo kintamojo y sumos įvetinius.

Imčiai $\mathfrak{s}_m$ išrinkti naudojamas dviejų fazių ėmimas:

$$\mathcal{U} \to \mathfrak{s}' \to \mathfrak{s}_m.$$

Pasinaudoję dviejų fazių sumos π^* įvertinio (1.8) apibrėžimu tyrimo kintamojo y sumai vertinti galime naudoti nepaslinktąjį įvertinį,

$$\hat{t}_{ym} = \sum_{i \in \mathfrak{s}_m} \frac{y_i}{\pi'_i \pi_{i|\mathfrak{s}'}}. \tag{2.1}$$

Tokiame dviejų fazių ėmime, kai pirmoje fazėje iš nagrinėjamos populiacijos $\mathcal{U}$ renkama imtis $\mathfrak{s}'$, o antroje fazėje iš turimos imties $\mathfrak{s}'$ renkama imtis $\mathfrak{s}_m$, kaip papildoma informacija, sudarant santykinį sumos įvertinį gali būti naudojamos pirmosios fazės imties $\mathfrak{s}'$ tyrimo kintamojo y reikšmės x_i.

Naudojant pirmosios fazės imtį $\mathfrak{s}'$ ir antrosios fazės imtį $\mathfrak{s}_m$ formuojamas santykinis sumos įvertinys

$$\hat{t}_{ym}^{sant} = \hat{t}_{xn'} \frac{\hat{t}_{ym}}{\hat{t}_{xm}} = \hat{t}_{xn'} \hat{r}_m, \tag{2.2}$$

čia

$$\hat{t}_{xn'} = \sum_{i \in \mathfrak{s}'} \frac{x_i}{\pi'_i}, \quad \hat{t}_{ym} = \sum_{i \in \mathfrak{s}_m} \frac{y_i}{\pi'_i \pi_{i|\mathfrak{s}'}}, \quad \hat{t}_{xm} = \sum_{i \in \mathfrak{s}_m} \frac{x_i}{\pi'_i \pi_{i|\mathfrak{s}'}}, \quad \hat{r}_m = \frac{\hat{t}_{ym}}{\hat{t}_{xm}}.$$

Santykinis sumos įvertinys t_{ym}^{sant} nėra nepaslinktasis. Jei tyrimo kintamojo y reikšmės y_i ir x_i pakankamai stipriai koreliuoja, tai, kaip rodo teorija, santykinis įvertinys $\hat{t}_{ym}^{sant}$ gali turėti mažesnę dispersiją negu tiesioginis, imties planu pagrįstas, sumos įvertinys $\hat{t}_{ym}$.

2.1 teiginys. *Santykinio sumos t_y įvertinio $\hat{t}_{ym}^{sant}$* (2.2) *dispersijai galime taikyti tokią apytikslės dispersijos išraišką*

$$\begin{aligned} \mathrm{AD}(\hat{t}_{ym}^{sant}) &= \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} \left(\pi'_{ij} - \pi'_i \pi'_j\right) \frac{y_i}{\pi'_i} \frac{y_j}{\pi'_j} \\ &\quad + \mathrm{E} \sum_{i,j \in \mathfrak{s}'} \left(\pi_{ij|\mathfrak{s}'} - \pi_{i|\mathfrak{s}'} \pi_{j|s'}\right) \frac{y_i - \hat{r}_{n'} x_i}{\pi'_i \pi_{i|\mathfrak{s}'}} \frac{y_j - \hat{r}_{n'} x_j}{\pi'_j \pi_{j|s'}}. \end{aligned} \tag{2.3}$$

Įrodymas. Dviejų fazių imties atveju santykinio sumos įvertinio $\hat{t}_{ym}^{sant}$ dispersijos forma gali būti užrašyta (1.1 pastaba)

$$\mathrm{D}(\hat{t}_{ym}^{sant}) = \mathrm{DE}(\hat{t}_{ym}^{sant} \mid \mathfrak{s}') + \mathrm{ED}(\hat{t}_{ym}^{sant} \mid \mathfrak{s}'),$$

Laikykime, kad $\hat{r}_m$ yra apytiksliai nepaslinktasis santykio

$$\hat{r}_{n'} = \frac{\hat{t}_{yn'}}{\hat{t}_{xn'}}$$

įvertinys antrosios fazės imties plano tikimybių atžvilgiu, esant fiksuotai pirmosios fazės imčiai $\mathfrak{s}'$:

$$\mathrm{E}(\hat{r}_m \mid s') \approx \hat{r}_{n'}, \quad tuomet \quad \mathrm{E}(\hat{t}_{ym}^{sant} \mid s') \approx \hat{t}_{xn'}\hat{r}_{n'} = \hat{t}_{yn'}.$$

Pasinaudoję šia apytiksle išraiška, gauname, kad

$$\mathrm{DE}(\hat{t}_{ym}^{sant} \mid \mathfrak{s}') \approx \mathrm{D}(\hat{t}_y) = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}} \left(\pi'_{ij} - \pi'_i\pi'_j\right)\frac{y_i}{\pi'_i}\frac{y_j}{\pi'_j}.$$

Nagrinėkime dispersijos antrąjį dėmenį $\mathrm{ED}(\hat{t}_{ym}^{sant} \mid \mathfrak{s}')$. Kadangi konstruojant santykinį sumos įvertinį $\hat{t}_{ym}^{sant}$ į jį įtraukiamas dviejų nepaslinktųjų sumų įvertinių santykis $\hat{r}_m$, tai tikslios ir tinkamos skaičiavimui dispersijos išraiškos nėra. Įvertinio $\hat{t}_{ym}^{sant}$ dispersiją $\mathrm{D}(\hat{t}_{ym}^{sant})$ keičiame apytiksle dispersijos išraiška

$$\mathrm{AD}(\hat{t}_{ym}^{sant}) = \sum_{i,j\in\mathfrak{s}'} \left(\pi_{ij|\mathfrak{s}'} - \pi_{i|\mathfrak{s}'}\pi_{j|\mathfrak{s}'}\right)\frac{y_i - \hat{r}_{n'}x_i}{\pi'_i\pi_{i|\mathfrak{s}'}}\frac{y_j - \hat{r}_{n'}x_j}{\pi'_j\pi_{j|\mathfrak{s}'}}.$$

Ši apytikslės dispersijos išraiška gaunama taikant Teiloro ištiesinimo metodą, kuris yra aptašytas 1.8.1 skyrelyje. □

2.1 Pastaba. Praktikoje galime taikyti šį dispersijos $\mathrm{D}(\hat{t}_{ym}^{sant})$ įvertinį:

$$\begin{aligned}\widehat{\mathrm{D}}(\hat{t}_{ym}^{sant}) &= \sum_{i,j\in\mathfrak{s}_m} \frac{\pi'_{ij} - \pi'_i\pi'_j}{\pi'_{ij}\pi_{ij|\mathfrak{s}'}}\frac{y_i}{\pi'_i}\frac{y_j}{\pi'_j} \\ &+ \sum_{i,j\in\mathfrak{s}_m} \frac{\pi_{ij|\mathfrak{s}'} - \pi_{i|\mathfrak{s}'}\pi_{j|\mathfrak{s}'}}{\pi_{ij|\mathfrak{s}'}}\frac{y_i - \hat{r}_m x_i}{\pi'_i\pi_{i|\mathfrak{s}'}}\frac{y_j - \hat{r}_m x_j}{\pi'_j\pi_{j|\mathfrak{s}'}}.\end{aligned}$$

Nežinomas parametras $\hat{r}_{n'}$ apytikslės dispersijos $\mathrm{AD}(\hat{t}_{ym}^{sant})$ išraiškoje keičiamas jo įverčiu $\hat{r}_m$.

Imčiai $\mathfrak{s}_u$ išrinkti taip pat naudojamas dviejų fazių ėmimas

$$\mathcal{U} \to \mathfrak{s}'^c = \mathcal{U} \setminus s' \to \mathfrak{s}_u.$$

Šiuo atveju pirmoje fazėje aibės $\mathfrak{s}'^c$ elementai nėra stebimi. Todėl naudojant imties $\mathfrak{s}_u$ duomenis tyrimo kintamojo y sumai vertinti taikome nepaslinkąjį populiacijos sumos t_y įvertinį dviejų fazių imties atveju (Särndal *et al* 1992):

$$\hat{t}_{yu} = \sum_{i \in \mathfrak{s}_u} \frac{y_i}{\pi_i'^c \pi_{i|\mathfrak{s}'^c}}, \tag{2.4}$$

čia $\pi_i'^c = 1 - \pi_i'$.

Remiantis 1.6 teiginyje pateikto nepaslinktojo sumos įvertinio dviejų fazių imties atveju dispersijos ir jos įvertinio išraiškomis, nagrinėjamo sumos įvertinio $\hat{t}_{yu}$ dispersija yra

$$\begin{aligned} \mathrm{D}(\hat{t}_{yu}) &= \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} \left(\pi_{ij}'^c - \pi_i'^c \pi_j'^c\right) \frac{y_i}{\pi_i'^c} \frac{y_j}{\pi_j'^c} \\ &+ \mathrm{E} \sum_{i,j \in \mathfrak{s}'^c} \left(\pi_{ij|\mathfrak{s}'^c} - \pi_{i|\mathfrak{s}'^c} \pi_{j|\mathfrak{s}'^c}\right) \frac{y_i}{\pi_i'^c \pi_{i|\mathfrak{s}'^c}} \frac{y_j}{\pi_j'^c \pi_{j|\mathfrak{s}'^c}}. \end{aligned} \tag{2.5}$$

Įvertinio $\hat{t}_{yu}$ dispersijai $\mathrm{D}(\hat{t}_{yu})$ vertinti galime naudoti įvertinį

$$\begin{aligned} \widehat{\mathrm{D}}(\hat{t}_{yu}) &= \sum_{i,j \in \mathfrak{s}_u} \frac{\pi_{ij}'^c - \pi_i'^c \pi_j'^c}{\pi_{ij}'^c \pi_{ij|\mathfrak{s}'^c}} \frac{y_i}{\pi_i'^c} \frac{y_j}{\pi_j'^c} \\ &+ \sum_{i,j \in \mathfrak{s}_u} \frac{\pi_{ij|\mathfrak{s}'^c} - \pi_{i|\mathfrak{s}'^c} \pi_{j|\mathfrak{s}'^c}}{\pi_{ij|\mathfrak{s}'^c}} \frac{y_i}{\pi_i'^c \pi_{i|\mathfrak{s}'^c}} \frac{y_j}{\pi_j'^c \pi_{j|\mathfrak{s}'^c}}. \end{aligned} \tag{2.6}$$

Iš (2.2) ir (2.4) įvertinių sudaromas sudėtinis populiacijos sumos t_y įvertinys:

$$\hat{t}_y^* = \alpha \hat{t}_{ym}^{san} + (1 - \alpha) \hat{t}_{yu}, \tag{2.7}$$

čia α, $\alpha \in (0{,}1)$ – laisvai pasirenkama konstanta.

2.2 teiginys. *Sudėtinio įvertinio* $\hat{t}_y^*$ (2.7) *apytikslę dispersiją galime išreikšti*

$$\mathrm{AD}(\hat{t}_y^*) = \alpha^2 \mathrm{AD}(\hat{t}_{ym}^{sant}) + (1-\alpha)^2 \mathrm{D}(\hat{t}_{yu}) + 2\alpha(1-\alpha)\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}). \tag{2.8}$$

Įrodymas. Naudodamiesi atsitiktinio dydžio dispersijos savybėmis, sudėtinio sumos įvertinio $\hat{t}_y^*$ (2.7) dispersiją užrašome:

$$\begin{aligned} \mathrm{D}(\hat{t}_y^*) &= \mathrm{D}(\alpha \hat{t}_{ym}^{san} + (1 - \alpha) \hat{t}_{yu}) \\ &\cong \alpha^2 \mathrm{D}(\hat{t}_{ym}^{sant}) + (1 - \alpha)^2 \mathrm{D}(\hat{t}_{yu}) + 2\alpha(1 - \alpha)\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}). \end{aligned}$$

Pakeitus santykinio sumos įvertinio $\hat{t}_{ym}^{san}$ dispersiją $\mathrm{D}(\hat{t}_{ym}^{sant})$ apytiksle dispersija $\mathrm{AD}(\hat{t}_{ym}^{sant})$ (2.3) gauname sudėtinio įvertinio $\hat{t}_y^*$ apytikslę dispersiją. □

2.2 Pastaba. Praktikoje galime taikyti šį dispersijos $\mathrm{D}(\hat{t}_y^*)$ įvertinį:

$$\widehat{\mathrm{D}}(\hat{t}_y^*) = \alpha^2\widehat{\mathrm{D}}(\hat{t}_{ym}^{sant}) + (1-\alpha)^2\widehat{\mathrm{D}}(\hat{t}_{yu}) + 2\alpha(1-\alpha)\widehat{\mathrm{Cov}}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}).$$

Nežinomus parametrus $\mathrm{AD}(\hat{t}_{ym}^{sant})$, $\mathrm{D}(\hat{t}_{yu})$ ir $\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu})$ keičiame atitinkamais jų įvertiniais $\widehat{\mathrm{D}}(\hat{t}_{ym}^{sant})$, $\widehat{\mathrm{D}}(\hat{t}_{yu})$ ir $\widehat{\mathrm{Cov}}(\hat{t}_{ym}^{sant}, \hat{t}_{yu})$.

Sudėtinis sumos įvertinys $\hat{t}_y^*$ (2.7) ir jo dispersijos įvertinys tiesiogiai priklauso nuo laisvai pasirenkamos konstantos α. Kitame teiginyje pateiksime α optimizavimo formulę α_{opt}, kurios pagalba galėsime minimizuoti įvertinio (2.7) apytikslę dispersiją (2.8).

2.3 teiginys. *α optimizavimo formulė α_{opt} yra*

$$\alpha_{opt} = \frac{\mathrm{D}_2 - \mathrm{C}}{\mathrm{D}_1 + \mathrm{D}_2 - 2\mathrm{C}}, \tag{2.9}$$

čia $\mathrm{D}_1 = \mathrm{AD}(\hat{t}_{ym}^{sant})$, $\mathrm{D}_2 = \mathrm{D}(\hat{t}_{yu})$ *ir* $\mathrm{C} = \mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu})$.

Įrodymas. Įvedame naujus pažymėjimus:

$$\mathrm{D}_1 = \mathrm{AD}(\hat{t}_{ym}^{sant}), \quad \mathrm{D}_2 = \mathrm{D}(\hat{t}_{yu}), \quad \mathrm{C} = \mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}).$$

Sudėtinio įvertinio $\hat{t}_y^*$ (2.7) apytikslę dispersiją (2.8) perrašome

$$\mathrm{AD}(\hat{t}_y^*) = \alpha^2\mathrm{D}_1 + (1-\alpha)^2\mathrm{D}_2 + 2\alpha(1-\alpha)\mathrm{C}.$$

Sprendžiame lygtį

$$\frac{\partial(\mathrm{AD}(\hat{t}_y^*)}{\partial\alpha} = 0,$$

koeficiento α atžvilgiu. Jos sprendinys yra ieškomas α_{opt}. □

Koeficientą α_{opt} vertiname:

$$\widehat{\alpha}_{opt} = \frac{\widehat{\mathrm{D}}_2 - \widehat{\mathrm{C}}}{\widehat{\mathrm{D}}_1 + \widehat{\mathrm{D}}_2 - 2\widehat{\mathrm{C}}},$$

sudėtinio įvertinio (2.7) išraiškoje koeficientą α keičiame optimalaus koeficiento α_{opt} įvertiniu $\hat{\alpha}_{opt}$. Taip surandamas optimalus sudėtinis sumos t_y įvertinys

$$\hat{t}^*_{y\ opt} = \widehat{\alpha}_{opt}\hat{t}^{san}_{ym} + (1-\widehat{\alpha}_{opt})\hat{t}_{yu}. \tag{2.10}$$

Su koeficientu α_{opt} gauto (2.7) įvertinio $\hat{t}^*_{y\ opt}$ apytikslė dispersija turi pavidalą

$$\mathrm{AD}(\hat{t}^*_{y\ opt}) = \frac{\mathrm{D}_1\mathrm{D}_2 - \mathrm{C}^2}{\mathrm{D}_1 + \mathrm{D}_2 - 2\mathrm{C}}, \tag{2.11}$$

jos įvertinys galėtų būti

$$\widehat{\mathrm{D}}(\hat{t}^*_{y\ opt}) = \frac{\widehat{\mathrm{D}}_1\widehat{\mathrm{D}}_2 - \widehat{\mathrm{C}}^2}{\widehat{\mathrm{D}}_1 + \widehat{\mathrm{D}}_2 - 2\widehat{\mathrm{C}}}.$$

Sekančiame poskyryje pateiksime sudarytų sumos t_y įvertinių bei jų dispersijų išraiškas, kai kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys.

2.1.1. Atskiras atvejis: paprastoji atsitiktinė negrąžintinė imtis

Tarkime n' dydžio paprastoji atsitiktinė imtis $\mathfrak{s}'$ renkama iš baigtinės N dydžio populiacijos $\mathcal{U}$, papildinys $\mathfrak{s}'^{c}$ taip pat yra paprastoji atsitiktinė imtis. m dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_m$ renkama iš n' dydžio imties $\mathfrak{s}'$ ir u dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_u$ renkama iš paprastosios atsitiktinės $N-n'$ dydžio imties $\mathfrak{s}'^{c}$. Tuomet π reikšmės apskaičiuojamos taip:

$$\pi'_i = \frac{n'}{N}, \qquad \pi'_{ij} = \frac{n'}{N}\frac{n'-1}{N-1}, \qquad \pi_{i|\mathfrak{s}'} = \frac{m}{n'}, \qquad \pi_{ij|\mathfrak{s}'} = \frac{m(m-1)}{n'(n'-1)},$$
$$\pi_{i|\mathfrak{s}'^c} = \frac{u}{N-n'}, \qquad \pi_{ij|\mathfrak{s}'^c} = \frac{u(u-1)}{(N-n')(N-n'-1)}, \qquad \pi'^c_i = \frac{N-n'}{N},$$

visiems $i,j \in \mathcal{U}$.

Baigtinės populiacijos tyrimo kintamojo y sumos t_y sudėtinį sanykinį įvertinį apibrėžiame taip:

$$\hat{t}^*_y = \alpha\hat{t}^{sant}_{ym} + (1-\alpha)\hat{t}_{yu}, \tag{2.12}$$

čia

$$\hat{t}^{sant}_{ym} = \frac{n'}{N}t_{xn'}\frac{\sum_{i\in\mathfrak{s}_m} y_i}{\sum_{i\in\mathfrak{s}_m} x_i}, \quad t_{xn'} = \sum_{i\in\mathfrak{s}'} x_i, \quad \hat{t}_{yu} = \frac{u}{N}\sum_{i\in\widehat{\mathfrak{s}}_u} y_i.$$

2.4 teiginys. *Populiacijos sumos t_y sudėtinio įvertinio $\hat{t}_y^*$ (2.12) apytikslė dispersija gali būti išreikšta*

$$\mathrm{AD}(\hat{t}_y^*) = \alpha^2 \mathrm{AD}(\hat{t}_{ym}^{sant}) + (1-\alpha)^2 \mathrm{D}(\hat{t}_{yu}) + 2\alpha(1-\alpha)\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}). \quad (2.13)$$

Čia naudojami žymenys:

Dviejų fazių imties atveju santykinio įvertinio $\hat{t}_{ym}^{sant}$ apytikslė dispersija

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

čia

$$s_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2, \qquad \mu_y = t_y/N,$$

$$s_r^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - r x_i)^2, \qquad r = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} x_i.$$

Nepaslinktojo sumos įvertinio $\hat{t}_{yu}$ dispersija dviejų fazių imties plano atveju

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

Įrodymas. Išraiškose (2.3) ir (2.6) π reikšmės keičiamos apskaičiuotomis paprastosios atsitiktinės imties plano atveju. □

Sekančiame teiginyje pateiksime dviejų sumos įvertinių $\hat{t}_{ym}^{sant}$ ir $\hat{t}_{yu}$ kovariacijos $\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu})$ išraišką, kuri yra apytikslės dispersijos $\mathrm{AD}(\hat{t}_y^*)$ (2.13) išraiškos narys.

2.5 teiginys. *Jei kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys, tai sumos įvertinių $\hat{t}_{ym}^{sant}$ ir $\hat{t}_{yu}$ kovariacija yra*

$$\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}) \cong -N s_y^2. \quad (2.14)$$

Įrodymas. Nagrinėjame dviejų sumos įvertinių $\hat{t}_{ym}^{sant}$ ir $\hat{t}_{yu}$ kovariaciją

$$\mathrm{Cov}(\hat{t}_{ym}^{sant}, \hat{t}_{yu}) = \mathrm{Cov}\Big(\frac{N}{n'} t_{xn'} \hat{r}_m, \hat{t}_{yu}\Big).$$

Kadangi santykio $r_{n'}$ įvertinys $\hat{r}_m$ nėra nepaslinktasis, naudodamiesi jo skleidiniu Teiloro eilute taško $(\hat{t}_{xm}, \hat{t}_{ym}) = (t_{xn'}, t_{yn'})$ aplinkoje ir imdami tik pirmosios eilės narius surandame apytikslę jo išraišką

$$\hat{r}_m \approx r_{n'} + \frac{1}{t_{xn'}}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm}).$$

Nagrinėjame kiekvieną kovariacijos

$$\mathrm{Cov}\left(\frac{N}{n'}(r_{n'}t_{xn'} + (\hat{t}_{ym} - r_{n'}\hat{t}_{xm})), \hat{t}_{yu}\right) =$$
$$= \mathrm{Cov}\left(\frac{N}{n'}t_{yn'}, \hat{t}_{yu}\right) + \mathrm{Cov}\left(\frac{N}{n'}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm}), \hat{t}_{yu}\right)$$

narį atskirai.

Pirmasis narys:

$$\mathrm{Cov}\left(\frac{N}{n'}t_{yn'}, \hat{t}_{yu}\right) = \mathrm{E}\left(\frac{N}{n'}t_{yn'} - \mathrm{E}\frac{N}{n'}t_{yn'}\right)\mathrm{E}(\hat{t}_{yu} - \mathrm{E}\hat{t}_{yu}).$$

Naudojantis atsitiktinių dydžių sąlyginių vidurkių savybėmis užrašome

$$\mathrm{E}\left(\mathrm{E}\left(\left(\frac{N}{n'}t_{yn'} - t_y\right)(\hat{t}_u - t_y) \mid \mathfrak{s}'\right)\right) =$$
$$= \mathrm{E}\left(\left(\frac{N}{n'}t_{yn'} - t_y\right)\left(\frac{N}{N-n'}\hat{t}_{yN-n'} - t_y\right)\right).$$

Apibrėžiame priklausymo imčiai $\mathfrak{s}'$ indikatorinį kintamąjį

$$\mathrm{I}_i = \begin{cases} 1, & \text{jeigu } i \in \mathfrak{s}'_n, \\ 0, & \text{jeigu } i \notin \mathfrak{s}'_n, \end{cases}$$

Nepriklausymo imčiai $\mathfrak{s}'$ tikimybė yra ekvivalenti imties $\mathfrak{s}'^c$ tikimybei $(1 - \mathrm{I}_i)$.

Įrašę jas į nagrinėjmą kovariacijos išraišką gauname

$$\begin{aligned}
&\mathrm{Cov}\left(\frac{N}{n'}t_{yn'},\hat{t}_{yu}\right) = \\
&= \mathrm{E}\left(\left(\frac{N}{n'}\sum_{i=1}^{N}\mathrm{I}_i y_i - t_y\right)\left(\frac{N}{N-n'}\sum_{j=1}^{N}(1-\mathrm{I}_j)y_j - t_y\right)\right) \\
&= \mathrm{E}\frac{N}{n'}\frac{N}{N-n'}\sum_{i=1}^{N}\mathrm{I}_i y_i\sum_{j=1}^{N}y_j(1-\mathrm{I_j}) - t_y^2 \\
&= \frac{N}{n'}\frac{N}{N-n'}\sum_{i,j=1}^{N}y_i y_j(\mathrm{EI}_i - \mathrm{EI}_i\mathrm{I}_j) - t_y^2.
\end{aligned}$$

Jei elementų indeksai i ir j sutampa, tuomet $\mathrm{EI}_i - \mathrm{EI}_i\mathrm{I}_j = 0$. Jei indeksai i ir j nesutampa, tuomet vidurkiai atitinkamai apskaičiuojami

$$\mathrm{EI}_i = \mathrm{P}(\mathfrak{s}' : i \in \mathfrak{s}') = \pi_i = \frac{n'}{N},$$

$$\begin{aligned}
\mathrm{EI}_i\mathrm{I}_j = \mathrm{P}(\mathfrak{s}' : i \in \mathfrak{s}', j \in \mathfrak{s}') &= \mathrm{P}(\mathfrak{s}' : i \in \mathfrak{s}' \mid j \in \mathfrak{s}')\mathrm{P}(j \in \mathfrak{s}') \\
&= \pi_{ij} = \frac{n'}{N}\frac{n'-1}{N-1}.
\end{aligned}$$

Apskaičiuotas vidurkių EI_i ir $\mathrm{EI}_i\mathrm{I}_j$ reikšmes rašome į kovariacijos išraišką ir sutvarkę reiškinį gauname

$$\mathrm{Cov}\left(\frac{N}{n'}t_{yn'},\hat{t}_{yu}\right) = -Ns_y^2,$$

čia $s_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^2$.
Antrasis narys:

$$\begin{aligned}
&\mathrm{Cov}\left(\frac{N}{n'}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm}),\hat{t}_{yu}\right) = \\
= \;&\mathrm{E}\left(\frac{N}{n'}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm}) - \mathrm{E}\frac{N}{n'}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm})\right)\mathrm{E}(\hat{t}_{yu} - \mathrm{E}\hat{t}_{yu}).
\end{aligned}$$

Pritaikome salyginių vidurkių savybes, apibrėžiame indikatorinius kintamuosius elementams priklausyti atitinkamoms imtims gauname

$$\mathrm{Cov}\left(\frac{N}{n'}(\hat{t}_{ym} - r_{n'}\hat{t}_{xm}), \hat{t}_{yu}\right) = 0.$$

□

2.3 Pastaba. Vertinant apytikslę dispersiją (2.13), naryje $\mathrm{AD}(\hat{t}_{ym}^{san})$ kintamojo y populiacijos dispersijos s_y^2 ir s_r^2 keičiamos į

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

$$\hat{s}_r^2 = \frac{1}{m-1}\sum_{i\in\mathfrak{s}_m}(y_i - \hat{r}x_i)^2, \quad \hat{r} = \frac{\sum_{i\in\mathfrak{s}_m} y_i}{\sum_{i\in\mathfrak{s}_m} x_i}.$$

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

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

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

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

2.1.2. Įvertinių tikslumo tyrimas, atliekant matematinį modeliavimą

Modeliavimui naudojami Statistikos departamento 2009 m. 1-ojo ir 2-ojo ketvirčių gyventojų užimtumo statistinio tyrimo duomenys. Abiejų ketvirčių imtims priklausantys elementai laikomi tyrimo populiacija, kurios dydis $N = 2425$. Modeliavimui naudojami imties dydžiai $n' = n = 200$, $m = u = 100$. Imties išrinkimas kartojamas $M = 1000$ kartų. Kiekvienam iš įvertinių $\hat{\theta} = \hat{t}_y^*$ (su koeficientu $\alpha = 0{,}5$), $\hat{t}_{y\ opt}^*$ (su optimaliu koeficientu α_{opt}) apskaičiuojamas empirinis įverčių vidurkis:

$$\bar{\hat{\theta}} = \frac{1}{M}\sum_{i=1}^{M}\hat{\theta},$$

poslinkis $\widehat{\text{Posl}}(\hat{\theta}) = \bar{\hat{\theta}} - t_y$, apytikslės dispersijos, vidutinės kvadratinė paklaidos įvertis $\widehat{\text{VKP}}(\hat{\theta}) = \text{AD}(\hat{\theta}) + (\widehat{\text{Posl}}(\hat{\theta}))^2$, santykinės vidutinės kvadratinės paklaidos įvertis

$$\widehat{\text{SVKP}}(\hat{\theta}) = \frac{\sqrt{\overline{\widehat{\text{VKP}}}(\hat{\theta})}}{\bar{\hat{\theta}}} \cdot (100\%), \quad \overline{\widehat{\text{VKP}}}(\hat{\theta}) = \frac{1}{M}\sum_{i=1}^{M} \widehat{\text{VKP}}(\hat{\theta}).$$

Apskaičiuojamas empirinis įverčių dispersijų vidurkis:

$$\overline{\widehat{\text{AD}}}(\hat{\theta}) = \frac{1}{M}\sum_{i=1}^{M} \widehat{\text{AD}}(\hat{\theta}),$$

Apskaičiuojamas empirinis įvertinio $\hat{\theta}$ variacijos koeficientas:

$$\hat{cv}(\hat{\theta}) = \frac{\sqrt{\overline{\widehat{\text{AD}}}(\hat{\theta})}}{\bar{\hat{\theta}}} \cdot 100(\%).$$

Palyginimui skaičiuojamas nepaslinktasis sumos įvertinys $\hat{t}_{y,PAI}$ (1.5) ir vertinama jo dispersija. Koeficientas α_{opt}, kuris yra įtrauktas į sudėtinį santykinį sumos įvertinį $\hat{t}^*_{y\ opt}$ (2.10) priklauso nuo kovariacijos, kuri kartais ignoruojama kaip nereikšminga. Nustatyti įtraukiamos kovariacijos įtaką įvertinio tikslumui taip pat lyginamas įvertinys (2.10) pakeičiant keficientą α_{opt}, koeficientu

$$\alpha_* = \frac{\widehat{D}_2}{\widehat{D}_1 + \widehat{D}_2}.$$

Tokį sudėtinį santykinį sumos įvertinį žymėsime $\hat{t}^*_{y\ *}$. Skaičiavimo rezultatai pateikiami 2.1 lentelėje.

2.1 lentelė. Reali populiacija: pagrindinės sumos įvertinių tikslumo charakteristikos (tikroji sumos reikšmė $t = 1\,022$)

Table 2.1. Real population: results of the estimation (true value of total $t = 1\,022$)

Įvertinys $\hat{\theta}$	Įverčių vidurkis	$\widehat{\text{Posl}}$	Tikrosios dispersijos	$\widehat{VKP}$	$\widehat{\text{SVKP}}$ (%)	$\bar{\hat{D}}$	$\hat{cv}$ (%)
$\hat{t}_{y,PAI}$	1 022,72	0,72	6 580,76	6 581,28	7,93	6 580,85	7,93
$\hat{t}^*_y$	1 025,49	3,49	4 956,84	4 969,02	6,87	4 964,81	6,87
$\hat{t}^*_{y\ *}$	1 023,85	1,85	4 746,23	4 748,08	6,73	4 750,15	6,73
$\hat{t}^*_{y\ opt}$	1 023,94	1,94	4 481,55	4 485,31	6,54	4 481,76	6,54

Iš modeliavimo rezultatų matyti, kad sudėtinio įvertinio $\hat{t}_y^*$, naudojančio papildomą informaciją dviejų fazių imties plano atveju, variacijos koeficiento įvertis sumažėjo, lyginant su nepaslinktuoju įvertiniu $\hat{t}_y$. Optimalaus koeficiento α_{opt} naudojimas sudėtiniame įvertinyje $\hat{t}_{y\ opt}^*$ (2.7) leidžia šio įvertinio kovariacijos koeficientą dar labiau sumažinti. Kovariacijos nario α_{opt} optimizavimo formulėje nepatartina ignoruoti, nors poslinkis šiek tiek sumažėja, tačiau dispersija, o tuo pačiu ir vidutinė kvadratinė paklaida padidėja.

Sekančiame poskyryje sudarysime santykinį sudėtinį sumos įvertinį papildomą informaciją naudodami visiems ketvirčiams, kuriems turime ankstesnių tyrimų duomenų.

2.2. Sumos vertinimas naudojant sudėtingesnę imties rinkimo schemą

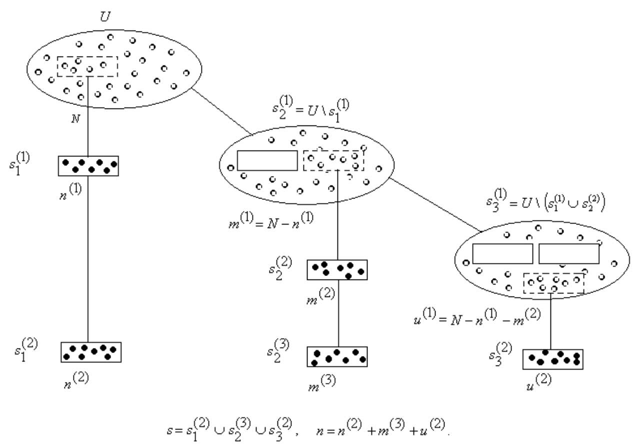
Nagrinėkime N dydžio baigtinę populiaciją $\mathcal{U} = \{1,\dots,\mathrm{N}\}$, kurios elementų tyrimo kintamojo y reikšmės duotu momentu yra $y_1,\dots,y_N$. Tarkime, kad šios populiacijos elementų skaičius ir patys elementai nekinta laikui bėgant. Imties rinkimo procedūra yra:

- **1 žingsnis**. Iš baigtinės populiacijos pagal tam tikrą imties planą renkama $n^{(1)}$ dydžio imtis $\mathfrak{s}_1^{(1)}$ su imties plano tikimybe $p(\mathfrak{s}_1^{(1)})$. Atitinkamai i-ojo ir j-ojo populiacijos elemento priklausymo šiai imčiai tikimybes žymėsime $\pi_{1i}^{(1)}$ ir $\pi_{1ij}^{(1)}$ visiems $i,j \in \mathcal{U}$.
- **2 žingsnis**. Pagal tam tikrą imties planą iš aibės $\mathfrak{s}_2^{(1)} = \mathcal{U}\setminus\mathfrak{s}_1^{(1)}$ elementų renkama $m^{(2)}$ dydžio imtis $\mathfrak{s}_2^{(2)}$ su imties plano tikimybe $p(\mathfrak{s}_2^{(2)}|\mathfrak{s}_2^{(1)})$. Atitinkamas i-ojo ir j-ojo populiacijos elemento priklausymo aibei $s_2^{(1)}$ tikimybes žymėsime $\pi_{2i}^{(1)}$ ir $\pi_{2ij}^{(1)}$, o imčiai $\mathfrak{s}_2^{(2)}$ sąlygines tikimybes žymėsime $\pi_{2i|\mathfrak{s}_2^{(1)}}^{(2)}$ ir $\pi_{2ij|\mathfrak{s}_2^{(1)}}^{(2)}$ visiems $i,j \in \mathcal{U}$.
- **3 žingsnis**. $n^{(2)}$ dydžio imtis $\mathfrak{s}_1^{(2)}$ pagal tam tikrą imties planą renkama iš jau išrinktos imties $\mathfrak{s}_1^{(1)}$ su imties plano tikimybe $p(\mathfrak{s}_1^{(2)}|\mathfrak{s}_1^{(1)})$. Atitinkamai i-ojo ir j-ojo populiacijos elemento priklausymo šiai imčiai sąlygines tikimybes žymėsime $\pi_{1i|\mathfrak{s}_1^{(1)}}^{(2)}$ ir $\pi_{1ij|\mathfrak{s}_1^{(1)}}^{(2)}$ visiems $i,j \in \mathcal{U}$.

  Taip pat $m^{(3)}$ dydžio imtis $\mathfrak{s}_2^{(3)}$ pagal tam tikrą imties planą renkama iš jau išrinktos imties $\mathfrak{s}_2^{(2)}$ su imties plano tikimybe $p(\mathfrak{s}_2^{(3)}|\mathfrak{s}_2^{(2)})$. Atitinkamai i-ojo ir j-ojo populiacijos elemento priklausymo šiai imčiai sąlygines tikimybes žymėsime $\pi^{(3)}_{2i|\mathfrak{s}_2^{(2)}}$ ir $\pi^{(3)}_{2ij|\mathfrak{s}_2^{(2)}}$ visiems $i,j \in \mathcal{U}$.

Iš aibės $\mathfrak{s}_3^{(1)} = U \setminus (\mathfrak{s}_1^{(1)} \cup \mathfrak{s}_2^{(2)})$ elementų pagal tam tikrą imties planą renkama $u^{(2)}$ dydžio imtis $\mathfrak{s}_3^{(2)}$ su imties plano tikimybe $p(\mathfrak{s}_3^{(2)}|\mathfrak{s}_3^{(1)})$. Atitinkamas i-ojo ir j-ojo populiacijos elemento priklausymo aibei $s_3^{(1)}$ tikimybes žymėsime $\pi^{(1)}_{3i}$ ir $\pi^{(1)}_{3ij}$, o imčiai $s_3^{(2)}$ sąlygines tikimybes žymėsime $\pi^{(2)}_{3i|\mathfrak{s}_3^{(1)}}$ ir $\pi^{(2)}_{3ij|\mathfrak{s}_3^{(1)}}$ visiems $i,j \in \mathcal{U}$.

Imties rinkimo procedūros grafinė schema vaizduojama 2.3 paveiksle.



2.3 pav. Imties rinkimo procedūros schema
Fig. 2.3. Sampling scheme

Iš pateiktos imties rinkimo procedūros schemos 2.3 paveiksle matyti, kad galutinę tyrimo imtį $\mathfrak{s}$ sudaro trijų imčių, atitinkamai $\mathfrak{s}_1^{(2)}$, $\mathfrak{s}_2^{(3)}$ ir $\mathfrak{s}_3^{(2)}$, sąjunga. Toliau konstruosime tris atskirus sumos įvertinius naudojant atitinkamai imčių $\mathfrak{s}_1^{(2)}$, $\mathfrak{s}_2^{(3)}$ ir $\mathfrak{s}_3^{(2)}$ duomenis.

Kaip matyti iš 2.3 paveiksle pateiktos imties rinkimo schemos, imčiai $\mathfrak{s}_1^{(2)}$ išrinkti naudojamas dviejų fazių ėmimas:

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

Tokiame dviejų fazių ėmime, kai pirmoje fazėje iš turimos populiacijos renkama imtis $\mathfrak{s}_1^{(1)}$, antroje fazėje iš turimos imties renkama imtis $\mathfrak{s}_1^{(2)}$, kaip papildoma informacija gali būti naudojama pirmosios fazės imties tyrimo kintamojo reikšmės. Pažymėkime imties $\mathfrak{s}_1^{(1)}$ elementų tyrimo kintamojo x reikšmes x_i, o antroje fazėje jau pasikeitusį tą patį kintamąjį pažymėkime y, o jo reikšmes y_i.

Naudojant pirmosios fazės imties $\mathfrak{s}_1^{(1)}$ ir antrosios fazės imties $\mathfrak{s}_1^{(2)}$ duomenis galime sudaryti santykinį tyrimo kintamojo y sumos įvertinį

$$\hat{t}_{1y}^{(2)sant} = \hat{t}_{1x}^{(1)} \frac{\hat{t}_{1y}^{(2)}}{\hat{t}_{1x}^{(2)}} = \hat{t}_{1x}^{(1)} \hat{r}_1, \tag{2.15}$$

čia

$$\hat{t}_{1x}^{(1)} = \sum_{i \in \mathfrak{s}_1^{(1)}} \frac{x_i}{\pi_{1i}^{(1)}}, \quad \hat{t}_{1y}^{(2)} = \sum_{i \in \mathfrak{s}_1^{(2)}} \frac{y_i}{\pi_{1i}^{(1)} \pi_{1i|\mathfrak{s}_1^{(1)}}^{(2)}},$$
$$\hat{t}_{1x}^{(2)} = \sum_{i \in \mathfrak{s}_1^{(2)}} \frac{x_i}{\pi_{1i}^{(1)} \pi_{1i|\mathfrak{s}_1^{(1)}}^{(2)}}, \quad \hat{r}_1 = \frac{\sum_{i \in \mathfrak{s}_1^{(2)}} x_i}{\sum_{i \in \mathfrak{s}_1^{(2)}} y_i}.$$

2.6 teiginys. *Santykinio sumos t_y įvertinio $\hat{t}_{1y}^{(2)sant}$ (2.15) dispersijai galime taikyti tokią apytikslės dispersijos išraišką*

$$\mathrm{AD}(\hat{t}_{1y}^{(2)sant}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\pi_{1ij}^{(1)} - \pi_{1i}^{(1)} \pi_{1j}^{(1)} \right) \frac{y_i}{\pi_{1i}^{(1)}} \frac{y_j}{\pi_{1j}^{(1)}}$$
$$+ \mathrm{E} \sum_{i \in s_1^{(1)}} \sum_{j \in s_1^{(1)}} \left(\pi_{1ij|s_1^{(1)}}^{(2)} - \pi_{1i|s_1^{(1)}}^{(2)} \pi_{1j|s_1^{(1)}}^{(2)} \right) \frac{R_i}{\pi_{1i}^{(1)} \pi_{1i|s_1^{(1)}}^{(2)}} \frac{R_j}{\pi_{1j}^{(1)} \pi_{1j|s_1^{(1)}}^{(2)}},$$

čia $R_i = y_i - r x_i$, $r = \sum_{i \in \mathcal{U}} y_i / \sum_{i \in \mathcal{U}} x_i$.

Šio teiginio įrodymas yra analogiškas 2.1 teiginio įrodymui.

2.4 Pastaba. Praktikoje galime taikyti šį dispersijos $\mathrm{D}(\hat{t}_{1y}^{(2)sant})$ įvertinį

$$\widehat{\mathrm{D}}(\hat{t}_{1y}^{(2)sant}) = \sum_{i \in s_1^{(2)}} \sum_{j \in s_1^{(2)}} \frac{\pi_{1ij}^{(1)} - \pi_{1i}^{(1)} \pi_{1j}^{(1)}}{\pi_{1ij}^{(1)} \pi_{1ij|s_1^{(1)}}^{(2)}} \frac{y_i}{\pi_{1i}^{(1)}} \frac{y_j}{\pi_{1j}^{(1)}}$$
$$+ \sum_{i \in s_1^{(2)}} \sum_{j \in s_1^{(2)}} \frac{\pi_{1ij|s_1^{(1)}}^{(2)} - \pi_{1i|s_1^{(1)}}^{(2)} \pi_{1j|s_1^{(1)}}^{(2)}}{\pi_{1ij|s_1^{(1)}}^{(2)}} \frac{\hat{R}_i}{\pi_{1i}^{(1)} \pi_{1i|s_1^{(1)}}^{(2)}} \frac{\hat{R}_j}{\pi_{1j}^{(1)} \pi_{1j|s_1^{(1)}}^{(2)}}.$$

Dydis $\hat{R}_i$ yra apibrėžiamas atitinkamoje išraiškoje R_i nežinomą parametrą r pakeičiant jo įvertiniu $\hat{r}_1$.

Nagrinėkime imties $\mathfrak{s}_2^{(3)}$ rinkimą. Imčiai $\mathfrak{s}_2^{(3)}$ išrinkti naudojamas trijų fazių ėmimas.

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

Tokiame trijų fazių ėmime, kai antroje fazėje iš imties $\mathfrak{s}_2^{(1)}$ renkama imtis $\mathfrak{s}_2^{(2)}$, o trečioje fazėje iš turimos imties renkama imtis $\mathfrak{s}_2^{(3)}$, kaip papildoma informacija gali būti naudojama antrosios fazės imties tyrimo kintamojo reikšmės. Pažymėkime imties $\mathfrak{s}_2^{(2)}$ elementų tyrimo kintamojo x reikšmes x_i, o trečioje fazėjė jau pasikeitusį tą patį kintamąjį pažymėkime y, o jo reikšmes y_i.

Naudojant antrosios fazės imtį $\mathfrak{s}_2^{(2)}$ ir trečiosios fazės imtį $\mathfrak{s}_2^{(3)}$ galime sudaryti santykinį tyrimo kintamojo y sumos įvertinį

$$\hat{t}_{2y}^{(3)sant} = \hat{t}_{2x}^{(2)} \frac{\hat{t}_{2y}^{(3)}}{\hat{t}_{2x}^{(3)}} = \hat{t}_{2x}^{(2)} \hat{r}_2 \tag{2.16}$$

čia

$$\hat{t}_{2x}^{(2)} = \sum_{i \in \mathfrak{s}_2^{(2)}} \frac{x_i}{\pi_{2i}^{(1)} \pi_{2i|\mathfrak{s}_2^{(1)}}^{(2)}}, \qquad \hat{t}_{2y}^{(3)} = \sum_{i \in \mathfrak{s}_2^{(3)}} \frac{y_i}{\pi_{2i}^{(1)} \pi_{2i|\mathfrak{s}_2^{(1)}}^{(2)} \pi_{2i|\mathfrak{s}_2^{(2)}}^{(3)}},$$

$$\hat{t}_{2x}^{(3)} = \sum_{i \in \mathfrak{s}_1^{(2)}} \frac{x_i}{\pi_{2i}^{(1)} \pi_{2i|\mathfrak{s}_2^{(1)}}^{(2)} \pi_{2i|\mathfrak{s}_2^{(2)}}^{(3)}}, \qquad \hat{r}_2 = \frac{\sum_{i \in \mathfrak{s}_2^{(3)}} x_i}{\sum_{i \in \mathfrak{s}_2^{(3)}} y_i}.$$

2.7 teiginys. *Santykinio sumos t_y įvertinio $\hat{t}_{ym}^{sant}$ (2.16) apytikslę dispersiją trijų fazių imties atveju galime išreikšti taip:*

$$\mathrm{AD}\hat{t}_{2y}^{(3)sant} = \mathrm{D}^{(1)} + \mathrm{D}^{(2)} + \mathrm{D}^{(3)}, \tag{2.17}$$

čia

$$\mathrm{D}^{(1)} = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\pi_{2ij}^{(1)c} - \pi_{2i}^{(1)c} \pi_{2j}^{(1)c} \right) \frac{y_i}{\pi_{2i}^{(1)c}} \frac{y_j}{\pi_{2j}^{(1)c}},$$

$$\begin{aligned}
\mathrm{D}^{(2)} &= \mathrm{E} \sum_{i \in s_2^{(1)c}} \sum_{j \in s_2^{(1)c}} \Big(\pi^{(2)}_{2ij|s_2^{(1)c}} - \pi^{(2)}_{2i|s_2^{(1)c}} \pi^{(2)}_{2j|s_2^{(1)c}} \Big) \\
&\quad \times \frac{y_i}{\pi^{(1)c}_{2i} \pi^{(2)}_{2i|s_2^{(1)c}}} \frac{y_j}{\pi^{(1)c}_{2j} \pi^{(2)}_{2j|s_2^{(1)c}}},
\end{aligned}$$

$$\begin{aligned}
\mathrm{D}^{(3)} &= \mathrm{E} \sum_{i \in s_2^{(2)}} \sum_{j \in s_2^{(2)}} \Big(\pi^{(3)}_{2ij|s_2^{(2)}} - \pi^{(3)}_{2i|s_2^{(2)}} \pi^{(3)}_{2j|s_2^{(2)}} \Big) \\
&\quad \times \frac{y_i - r x_i}{\pi^{(1)c}_{2i} \pi^{(2)}_{2i|s_2^{(1)c}} \pi^{(3)}_{2i|s_2^{(2)}}} \frac{y_j - r x_j}{\pi^{(1)c}_{2j} \pi^{(2)}_{2j|s_2^{(1)c}} \pi^{(3)}_{2j|s_2^{(2)}}}.
\end{aligned}$$

Įrodymas. Trijų fazių imties atveju santykinio sumos įvertinio $\widehat{t}_{2y}^{(3)sant}$ dispersijos forma yra

$$\begin{aligned}
\mathrm{D}(\hat{t}_{2y}^{(3)sant}) &= \mathrm{DE}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(1)}) + \mathrm{ED}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(1)}) \\
&= \mathrm{D}\Big(\mathrm{E}\big(\mathrm{E}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)})\big) \mid \mathfrak{s}^{(1)} \Big) \\
&\quad + \mathrm{E}\bigg(\mathrm{D}\Big(\mathrm{E}\big(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}\big) \mid \mathfrak{s}^{(1)} \Big) \bigg) \\
&\quad + \mathrm{E}\bigg(\mathrm{E}\Big(\mathrm{D}\big(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}\big) \mid \mathfrak{s}^{(1)} \Big) \bigg).
\end{aligned}$$

Nagrinėkime į pirmąjį ir antrąjį narį įeinantį vidurkį $\mathrm{E}\big(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}\big)$. Laikykime, kad $\hat{r}^{(2)} = \hat{t}_{2y}^{(2)} / \hat{t}_{2x}^{(2)}$ yra apytiksliai nepaslinktasis santykio

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

įvertinys trečiosios fazės imties plano tikimybių atžvilgiu, esant fiksuotai antrosios fazės imčiai $\mathfrak{s}^{(2)}$:

$$\mathrm{E}\big(\hat{r}^{(3)} \mid \mathfrak{s}^{(2)}\big) \approx \hat{r}^{(2)}, \quad tuomet \quad \mathrm{E}\big(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}\big) \approx \hat{t}_{2x}^{(2)} \hat{r}^{(2)} = \hat{t}_{2y}^{(2)}.$$

Įrašę gautą išraišką į pirmus du narius, toliau nagrinėjame išraišką

$$\mathrm{DE}(\hat{t}_{2y}^{(2)} \mid \mathfrak{s}^{(1)}) + \mathrm{ED}(\hat{t}_{2y}^{(2)} \mid \mathfrak{s}^{(1)}),$$

kuri yra lygi nepaslinktojo sumos įvertinio $\hat{t}_{2y}^{(2)}$ dviejų fazių imties atveju dispersijos formai. Remiantis 1.6 teiginiu, galime užrašyti

$$\mathrm{DE}(\hat{t}_{2y}^{(2)} \mid \mathfrak{s}^{(1)}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\pi_{2ij}^{(1)c} - \pi_{2i}^{(1)c}\pi_{2j}^{(1)c}\right)\frac{y_i}{\pi_{2i}^{(1)c}}\frac{y_j}{\pi_{2j}^{(1)c}},$$

o

$$\begin{aligned}\mathrm{ED}(\hat{t}_{2y}^{(2)} \mid \mathfrak{s}^{(1)}) &= \mathrm{E}\sum_{i\in s_2^{(1)c}}\sum_{j\in s_2^{(1)c}}\left(\pi^{(2)}_{2ij|s_2^{(1)c}} - \pi^{(2)}_{2i|s_2^{(1)c}}\pi^{(2)}_{2j|s_2^{(1)c}}\right)\\ &\quad\times\frac{y_i}{\pi_{2i}^{(1)c}\pi^{(2)}_{2i|s_2^{(1)c}}}\frac{y_j}{\pi_{2j}^{(1)c}\pi^{(2)}_{2j|s_2^{(1)c}}}.\end{aligned}$$

Toliau nagrinėjame narį

$$\mathrm{E}\left(\mathrm{E}\left(\mathrm{D}\left(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}\right) \mid \mathfrak{s}^{(1)}\right)\right) = \mathrm{ED}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}).$$

Kadangi į sumos įvertinį $\hat{t}_{2y}^{(3)sant}$ yra įtrauktas dviejų nepaslinktųjų sumų santykis $\hat{r}^{(3)}$, tai tikslios dispersijos šis įvertinys neturi. Įvertinio $\hat{t}_{2y}^{(3)sant}$ dispersija $\mathrm{D}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)})$ keičiama apytiksle dispersijos išaiška, kuri gaunama taikant Teiloro ištiesinimo metodą

$$\begin{aligned}\mathrm{EAD}(\hat{t}_{2y}^{(3)sant} \mid \mathfrak{s}^{(2)}) &= \mathrm{E}\sum_{i\in s_2^{(2)}}\sum_{j\in s_2^{(2)}}(\pi^{(3)}_{2ij|s_2^{(2)}} - \pi^{(3)}_{2i|s_2^{(2)}}\pi^{(3)}_{2j|s_2^{(2)}})\\ &\quad\times\frac{y_i - \hat{r}^{(2)}x_i}{\pi_{2i}^{(1)c}\pi^{(2)}_{2i|s_2^{(1)c}}\pi^{(3)}_{2i|s_2^{(2)}}}\frac{y_j - \hat{r}^{(2)}x_j}{\pi_{2j}^{(1)c}\pi^{(2)}_{2j|s_2^{(1)c}}\pi^{(3)}_{2j|s_2^{(2)}}}.\end{aligned}$$

□

Imčiai $\mathfrak{s}_3^{(2)}$ išrinkti, naudojamas dviejų fazių ėmimas:

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

Pirmoje fazėje imties $\mathfrak{s}_3^{(1)}$ elementai nėra stebimi, todėl tyrimo kintamojo y sumos vertinimui naudojame dviejų fazių sumos nepaslinktąjį įvertinį

$$\hat{t}_{3y}^{(2)} = \sum_{i \in \mathfrak{s}_3^{(2)}} \frac{y_i}{\pi_{3i}^{(1)} \pi_{3i|\mathfrak{s}_3^{(1)}}^{(2)}}. \tag{2.18}$$

Nepaslinktojo sumos įvertinio $\hat{t}_{3y}^{(2)}$ dviejų fazių imties atveju dispersija yra

$$\mathrm{D}(\hat{t}_{3y}^{(2)}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\pi_{3ij}^{(1)} - \pi_{3i}^{(1)} \pi_{3j}^{(1)} \right) \frac{y_i}{\pi_{3i}^{(1)}} \frac{y_j}{\pi_{3j}^{(1)}}$$
$$+ \mathrm{E} \sum_{i \in s_3^{(1)}} \sum_{j \in s_3^{(1)}} \left(\pi_{3ij|s_3^{(1)}}^{(2)} - \pi_{3i|s_3^{(1)}}^{(2)} \pi_{3j|s_3^{(1)}}^{(2)} \right) \frac{y_i}{\pi_{3i}^{(1)} \pi_{3i|s_3^{(1)}}^{(2)}} \frac{y_j}{\pi_{3j}^{(1)} \pi_{3j|s_3^{(1)}}^{(2)}}.$$

o dispersijai $\mathrm{D}(\hat{t}_{3y}^{(2)})$ vertinti galime naudoti įvertinį

$$\widehat{\mathrm{D}}(\hat{t}_{3y}^{(2)}) = \sum_{i \in s_3^{(2)}} \sum_{j \in s_3^{(2)}} \frac{\pi_{3ij}^{(1)} - \pi_{3i}^{(1)} \pi_{3j}^{(1)}}{\pi_{3ij}^{(1)} \pi_{3ij|s_3^{(1)}}^{(2)}} \frac{y_i}{\pi_{3i}^{(1)}} \frac{y_j}{\pi_{3j}^{(1)}}$$
$$+ \sum_{i \in s_3^{(2)}} \sum_{j \in s_3^{(2)}} \frac{\pi_{3ij|s_3^{(1)}}^{(2)} - \pi_{3i|s_3^{(1)}}^{(2)} \pi_{3j|s_3^{(1)}}^{(2)}}{\pi_{3ij|s_3^{(1)}}^{(2)}} \frac{y_i}{\pi_{3i}^{(1)} \pi_{3i|s_3^{(1)}}^{(2)}} \frac{y_j}{\pi_{3j}^{(1)} \pi_{3j|s_3^{(1)}}^{(2)}}.$$

Iš įvertinių (2.15), (2.16) ir (2.18) sudaromas sudėtinis populiacijos sumos t_y įvertinys

$$\hat{t}_y^{**} = \alpha \hat{t}_{1y}^{(2)sant} + \beta \hat{t}_{2y}^{(3)sant} + \gamma \hat{t}_{3y}^{(2)}, \tag{2.19}$$

kur α, β ir γ yra konstantos ($0 < \alpha < 1, 0 < \beta < 1, 0 < \gamma < 1$), bet to $\alpha + \beta + \gamma = 1$.

2.8 teiginys. *Sudėtinio sumos įvertinio* (2.19) *apytikslę dispersiją galime išreikšti taip:*

$$\begin{aligned} \mathrm{AD}(\hat{t}_y^{**}) \;\; = \;\; & \alpha^2 \mathrm{AD}(\hat{t}_{1y}^{(2)sant}) + \beta^2 \mathrm{AD}(\hat{t}_{2y}^{(3)sant}) + \gamma^2 \mathrm{D}(\hat{t}_{3y}^{(2)}) \\ & + 2\alpha\beta \mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{2y}^{(3)sant}) + 2\alpha\gamma \mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{3y}^{(2)}) \\ & + 2\beta\gamma \mathrm{cov}(\hat{t}_{2y}^{(3)sant}, \hat{t}_{3y}^{(2)}). \end{aligned} \tag{2.20}$$

Įrodymas. Naudodamiesi atsitiktinio dydžio dispersijos savybėmis sumos įvertinio $\hat{t}_y^{**}$ (2.19) dispersiją galime užrašyti:

$$\begin{aligned}
\mathrm{D}(\hat{t}_y^{**}) &= \mathrm{D}(\alpha\hat{t}_{1y}^{(2)sant} + \beta\hat{t}_{2y}^{(3)sant} + \gamma\hat{t}_{3y}^{(2)}) \\
&\cong \alpha^2\mathrm{D}(\hat{t}_{1y}^{(2)sant}) + \beta^2\mathrm{D}(\hat{t}_{2y}^{(3)sant}) + \gamma^2\mathrm{D}(\hat{t}_{3y}^{(2)}) \\
&\quad +2\alpha\beta\mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{2y}^{(3)sant}) + 2\alpha\gamma\mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{3y}^{(2)}) \\
&\quad +2\beta\gamma\mathrm{cov}(\hat{t}_{2y}^{(3)sant}, \hat{t}_{3y}^{(2)}).
\end{aligned}$$

Nežinomas santykinių sumos įvertinių, atitinkamai $\hat{t}_{1y}^{(2)sant}$ ir $\hat{t}_{2y}^{(3)sant}$, dispersijas keičiame apytikslėmis dispersijomis $\mathrm{AD}(\hat{t}_{1y}^{(2)sant})$ ir $\mathrm{AD}(\hat{t}_{2y}^{(3)sant})$ atitinkamai. □

2.2.1. Atskiras atvejis: paprastoji atsitiktinė negrąžintinė imtis

Tarkime $n^{(1)}$ dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_1^{(1)}$ renkama iš N dydžio baigtinės populiacijos $\mathcal{U}$. Iš turimos $\mathfrak{s}_1^{(1)}$ imties renkama $n^{(2)}$ dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_1^{(2)}$, ir $m^{(2)}$ dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_2^{(2)}$ renkama iš $m^{(1)} = N - n^{(1)}$ dydžio aibės $\mathfrak{s}_2^{(1)} = \mathcal{U}\backslash\mathfrak{s}_1^{(1)}$ elementų. $m^{(3)}$ dydžio paprastoji atsitiktinė imtis renkama iš išrinktos imties $\mathfrak{s}_2^{(2)}$ ir galiausiai $u^{(2)}$ dydžio paprastoji atsitiktinė imtis $\mathfrak{s}_3^{(2)}$ renkama iš $u^{(1)} = N - n^{(1)} - m^{(2)}$ dydžio aibės $\mathfrak{s}_3^{(1)} = \mathcal{U}\backslash(\mathfrak{s}_1^{(1)} \cup \mathfrak{s}_2^{(2)})$ elementų. Tuomet π reikšmės apskaičiuojamos taip:

$$\pi_{1i}^{(1)} = \frac{n^{(1)}}{N}, \quad \pi_{1ij}^{(1)} = \frac{n^{(1)}}{N}\frac{n^{(1)}-1}{N-1}, \quad \pi_{1i|\mathfrak{s}_1^{(1)}}^{(2)} = \frac{n^{(2)}}{n^{(1)}},$$

$$\pi_{1ij|\mathfrak{s}_1^{(1)}}^{(2)} = \frac{n^{(2)}}{n^{(1)}}\frac{n^{(2)}-1}{n^{(1)}-1}, \quad \pi_{2i}^{(1)} = \frac{N-n^{(1)}}{N}, \quad \pi_{2ij}^{(1)} = 1 - \frac{n^{(1)}}{N}\frac{n^{(1)}-1}{N-1},$$

$$\pi_{2i|\mathfrak{s}_2^{(1)}}^{(2)} = \frac{m^{(2)}}{N-n^{(1)}}, \quad \pi_{2ij|\mathfrak{s}_2^{(1)}}^{(2)} = \frac{m^{(2)}}{N-n^{(1)}}\frac{m^{(2)}-1}{N-n^{(1)}-1}, \quad \pi_{2i|\mathfrak{s}_2^{(2)}}^{(3)} = \frac{m^{(3)}}{m^{(2)}},$$

$$\pi_{2ij|\mathfrak{s}_2^{(2)}}^{(3)} = \frac{m^{(3)}}{m^{(2)}}\frac{m^{(3)}-1}{m^{(2)}-1}, \quad \pi_{3i}^{(1)} = \frac{N-n^{(1)}-m^{(2)}}{N},$$

$$\pi_{3ij}^{(1)} = \frac{N-n^{(1)}-m^{(2)}}{N}\frac{N-n^{(1)}-m^{(2)}-1}{N-1}, \quad \pi_{3i|\mathfrak{s}_3^{(1)}}^{(2)} = \frac{u^{(2)}}{N-n^{(1)}-m^{(2)}},$$

$$\pi^{(2)}_{3ij|\mathfrak{s}_3^{(1)}} = \frac{u^{(2)}}{N-n^{(1)}-m^{(2)}}\frac{u^{(2)}-1}{N-n^{(1)}-m^{(2)}-1},$$

visiems $i,j \in \mathcal{U}$.

Baigtinės populiacijos tyrimo kintmojo y sumos t_y sudėtinis įvertinys apibrėžiamas taip:

$$\hat{t}_y^{**} = \alpha\hat{t}_{1y}^{(2)sant} + \beta\hat{t}_{2y}^{(3)sant} + \gamma\hat{t}_{3y}^{(2)}, \tag{2.21}$$

čia

$$\hat{t}_{1y}^{(2)sant} = \frac{N}{n^{(1)}}t_{1x}^{(2)}\frac{\sum_{i\in\mathfrak{s}_1^{(2)}} y_i}{\sum_{i\in\mathfrak{s}_1^{(2)}} y_i}, \quad t_{1x}^{(2)} = \sum_{i\in\mathfrak{s}_1^{(1)}} x_i,$$

$$\hat{t}_{2y}^{(3)sant} = \frac{N}{m^{(2)}}t_{2x}^{(2)}\frac{\sum_{i\in\mathfrak{s}_2^{(3)}} y_i}{\sum_{i\in\mathfrak{s}_2^{(3)}} x_i}, \quad t_{2x}^{(2)} = \sum_{i\in\mathfrak{s}_2^{(2)}} x_i, \quad \hat{t}_{3y}^{(2)} = \frac{N}{u^{(2)}}\sum_{i\in\mathfrak{s}_3^{(2)}} y_i.$$

2.9 teiginys. *Populiacijos sumos t_y sudėtinio įvertinio t_y^{**} (2.21) apytikslė dispersija gali būti išreikšta*

$$\begin{aligned}\mathrm{AD}(\hat{t}_y^{**}) &= \alpha^2\mathrm{AD}(\hat{t}_{1y}^{(2)sant}) + \beta^2\mathrm{AD}(\hat{t}_{2y}^{(3)sant}) + \gamma^2\mathrm{D}(\hat{t}_{3y}^{(2)})\\ &+2\alpha\beta\mathrm{cov}(\hat{t}_{1y}^{(2)sant},\hat{t}_{2y}^{(3)sant}) + 2\alpha\gamma\mathrm{cov}(\hat{t}_{1y}^{(2)sant},\hat{t}_{3y}^{(2)})\\ &+2\beta\gamma\mathrm{cov}(\hat{t}_{2y}^{(3)sant},\hat{t}_{3y}^{(2)}).\end{aligned}$$

Čia naudojami žymenys:

Dviejų fazių imties atveju santykinio sumos įvertinio $\hat{t}_{1y}^{sant}$ apytikslė dispersija

$$\mathrm{AD}(\hat{t}_{1y}^{(2)sant}) = N^2\left(1-\frac{n^{(1)}}{N}\right)\frac{s_{1y}^2}{n^{(1)}} + N^{(2)}\left(1-\frac{n^{(2)}}{n^{(1)}}\right)\frac{s_{1r}^2}{n^{(2)}}, \tag{2.22}$$

$$s_{1y}^2 = \frac{1}{N-1}\sum_{i=1}^N(y_i-\mu_y)^2, \quad \mu_y = \frac{1}{N}\sum_{i=1}^N y_i,$$

$$s_{1r}^2 = \frac{1}{N-1}\sum_{i=1}^N(y_i-rx_i)^2, \quad r = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}.$$

Trijų fazių imties atveju santykinio sumos įvertinio $\hat{t}_{2y}^{sant}$ apytikslė dispersija

$$\mathrm{AD}(\hat{t}_{2y}^{(3)sant}) = N^2\left(1 - \frac{m^{(2)}}{N}\right)\frac{s_{2y}^2}{m^{(2)}} + N^2\left(1 - \frac{m^{(3)}}{m^{(2)}}\right)\frac{s_{2r}^2}{m^{(3)}}, \tag{2.23}$$

čia $s_{2y}^2 = s_{1y}^2$ ir $s_{2r}^2 = s_{1r}^2$ yra pateiktos (2.22) išraiškoje.

Dviejų fazių imties atveju nepaslinktojo sumos įvertinio $\hat{t}_{3y}$ dispersija

$$\mathrm{D}(\hat{t}_{3y}^{(2)}) = N^2\left(1 - \frac{u^{(2)}}{N}\right)\frac{s_{3y}^2}{u^{(2)}}, \tag{2.24}$$

čia $s_{3y}^2 = s_{1y}^2$ yra pateikta (2.22) išraiškoje. Atitinkamai $\operatorname{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{2y}^{(3)sant}) = \operatorname{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{3y}^{(3)}) = \operatorname{cov}(\hat{t}_{2y}^{(3)sant}, \hat{t}_{3y}^{(2)}) = -Ns_{1y}^2$.

Įrodymas. Išraiškose (2.16), (2.17) ir (2.19) π reikšmės keičiamos apskaičiuotomis paprastosios atsitiktinės imties plano atveju. Kovariacijų išraiškos gaunamos remiantis 2.5 teiginio įrodymu. □

2.5 Pastaba. Vertinant dispersiją $\mathrm{D}(\hat{t}_y^{**})$, naryje $\mathrm{AD}(\hat{t}_{1y}^{(2)sant})$ populiacijos tyrimo kintamojo y dispersijos s_{1y}^2 ir s_{1r}^2 keičiamos į

$$\hat{s}_{1y}^2 = \frac{1}{n^{(2)} - 1}\sum_{i\in s_1^{(2)}}(y_i - \overline{y}^{(2)})^2, \quad \overline{y}^{(2)} = \frac{1}{n^{(2)}}\sum_{i\in s_1^{(2)}} y_i,$$

ir

$$\hat{s}_{1r}^2 = \frac{1}{n^{(2)} - 1}\sum_{i\in s_1^{(2)}}\left(y_i - \hat{r}^{(2)}x_i\right)^2,$$

čia $\hat{r}^{(2)}$ pateikta (2.21) išraiškoje.

Naryje $\mathrm{AD}(\hat{t}_{2y}^{(3)sant})$ populiacijos tyrimo kintamojo y dispersijos s_{2y}^2 ir s_{2r}^2 keičiamos į

$$\hat{s}_{2y}^2 = \frac{1}{m^{(3)} - 1}\sum_{i\in s_2^{(3)}}(y_i - \overline{y}^{(3)})^2, \quad \overline{y}^{(3)} = \frac{1}{m^{(3)}}\sum_{i\in s_2^{(3)}} y_i,$$

ir

$$\hat{s}_{2r}^2 = \frac{1}{m^{(3)} - 1}\sum_{i\in s_2^{(3)}}\left(y_i - \hat{r}^{(3)}x_i\right)^2,$$

čia $\hat{r}^{(3)}$ pateikta (2.21) išraiškoje.

Naryje $\mathrm{D}(\hat{t}_{3y}^{(2)})$ populiacijos tyrimo kintamojo y populiacijos dispersija s_{3y}^2 keičiama į

$$\hat{s}_{3y}^2 = \frac{1}{u^{(2)}-1}\sum_{i\in s_3^{(2)}}(y_i-\overline{y})^2, \quad \overline{y} = \frac{1}{u^{(2)}}\sum_{i\in s_3^{(2)}} y_i,$$

nariuose $\mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{2y}^{(3)sant})$, $\mathrm{cov}(\hat{t}_{1y}^{(2)sant}, \hat{t}_{3y}^{(2)})$ ir $\mathrm{cov}(\hat{t}_{2y}^{(3)sant}, \hat{t}_{3y}^{(2)})$ kintamojo y populiacijos dispersija s_{1y}^2 keičiama atitinkamai į

$$\hat{s}_{12y}^2 = \frac{1}{n^{(2)}+m^{(3)}-1}\sum_{i\in(\mathfrak{s}_1^{(2)}\cup\mathfrak{s}_2^{(3)})}(y_i-\overline{y})^2, \quad \overline{y} = \frac{1}{n^{(2)}+m^{(3)}}\sum_{i\in(\mathfrak{s}_1^{(2)}\cup\mathfrak{s}_2^{(3)})} y_i,$$

$$\hat{s}_{13y}^2 = \frac{1}{n^{(2)}+u^{(2)}-1}\sum_{i\in(\mathfrak{s}_1^{(2)}\cup\mathfrak{s}_3^{(2)})}(y_i-\overline{y})^2, \quad \overline{y} = \frac{1}{n^{(2)}+u^{(2)}}\sum_{i\in(\mathfrak{s}_1^{(2)}\cup\mathfrak{s}_3^{(2)})} y_i,$$

ir

$$\hat{s}_{23y}^2 = \frac{1}{m^{(3)}+u^{(2)}-1}\sum_{i\in(\mathfrak{s}_2^{(3)}\cup\mathfrak{s}_3^{(2)})}(y_i-\overline{y})^2, \quad \overline{y} = \frac{1}{m^{(3)}+u^{(2)}}\sum_{i\in(\mathfrak{s}_2^{(3)}\cup\mathfrak{s}_3^{(2)})} y_i.$$

2.2.2. Matematinis modeliavimas: papildomos informacijos įtaka įvertinių tikslumui

Modeliavimui naudojami Statistikos departamento 2008 m. visų keturių ketvirčių ir 2009 m. 1-ojo ir 2-ojo ketvirčių gyventojų užimtumo statistinio tyrimo duomenys. Visų turimų ketvirčių imtims priklausantys elementai laikomi tyrimo populiacija, kurios dydis $N = 4485$. Modeliavimui visose fazėse renkamos paprastosios atsitiktinės imtys, naudojami imties dydžiai $n^{(1)} = 100$, $n^{(2)} = 50$, $m^{(2)} = 200$, $m^{(3)} = 100$, $u^{(2)} = 50$. Pakartojus imties išrinkimą ir įverčio $\hat{t}_y^{**}$ skaičiavimą $M = 1000$ kartų, skaičiuojamos tos pačios įvertinių tikslumo charakteristikos, kaip ir įvertiniams $\hat{t}_y^*$ ir $\hat{t}_{y\ opt}^*$ 2.1.2 skyrelyje. Skaičiavimo rezultatai pateikiami 2.2 lentelėje.

Iš modeliavimo rezultatų matome, kad sumos įverčių variacijos koeficientas taikant įvertinį $\hat{t}^{**}$ dar labiau sumažėjo lyginant su kitais įvertiniais, tačiau tai gaunama padidėjusio poslinkio sąskaita.

2.2 lentelė. Reali populiacija: pagrindinės sumos įvertinių tikslumo charakteristikos (tikroji sumos reikšmė $t = 1\,022$)

Table 2.2. Real population: results of the estimation (true value of total $t = 1\,022$)

Įvertinys $\hat{\theta}$	Įverčių vidurkis	$\widehat{\text{Posl}}$	Tikrosios dispersijos	$\widehat{VKP}$	$\widehat{\text{SVKP}}$ (%)	$\overline{\hat{D}}$	$\hat{cv}$ (%)
$\hat{t}_y$	1 022,72	0,72	6 580,76	6 581,28	7,93	6 580,85	7,93
$\hat{t}_y^*$	1 025,49	3,49	4 956,84	4 969,02	6,87	4 964,81	6,87
$\hat{t}_{y\,*}^*$	1 023,85	1,85	4 746,23	4 748,08	6,73	4 750,15	6,73
$\hat{t}_{y\,opt}^*$	1 023,94	1,94	4 481,55	4 485,31	6,54	4 481,76	6,54
$\hat{t}_y^{**}$	1 013,79	-8,21	4 144,89	4 212,33	6,40	4 319,53	6,48

2.3. Antrojo skyriaus apibendrinimas

1. Sudaryti baigtinės populiacijos sumos sudėtiniai santykiniai įvertiniai. Pasiūlytos tokių įvertinių apytikslių dispersijų išraiškos bet kokiam imties planui, taip pat atskiru, paprastosios atsitiktinės imties, atveju.

2. Sudarant baigtinės populiacijos sumos įvertinius esant imties rotacijai, žinoma papildoma informacija iš ankstesnių imties rinkimų turi būti išnaudojama, siekant pagerinti įvertinių tikslumą.

3. Atliekant modeliavimą su realiais duomenimis buvo parodyta papildomos informacijos, žinomos iš ankstesnių tyrimų, įtaka baigtinės populiacijos sumos vertinimo tikslumui rotuojant imtį.

3

Baigtinės populiacijos sudėtingesnių parametrų vertinimas esant imties rotacijai

Ankstesniame skyriuje nagrinėjome baigtinės populiacijos tyrimo kintamojo reikšmių sumos sudėtinio santykinio įvertinio sudarymo ypatybes, esant imties rotacijai. Sudėtingesnių parametrų vertinimas naudojant imties rotaciją nėra plačiai nagrinėjamas literatūroje. Šiame skyriuje sudarysime populiacijos tyrimo kintamojo sudėtinius pasiskirstymo funkcijos įvertinius.

3.1. Pasiskirstymo funkcijos vertinimas

3.1.1. Tradicinis pasiskirstymo funkcijos įvertinys

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{u_1, \ldots, u_N\}$, iš kurios išrinkta tikimybinė imtis $\mathfrak{s}$ pagal imties planą $p(\mathfrak{s})$. Tarkime, kad y yra populiacijoje $\mathcal{U}$ apibrėžtas tyrimo kintamasis, kurio reikšmės $\{y_1, y_2, \ldots y_N\}$. Dominantis parametras – baigtinės populiacijos tyrimo kintamojo y pasiskirstymo funkcija

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

čia su kiekvienu realiu skaičiumi z $(-\infty < z < \infty)$ apibrėžta aibė $A_z = \{l \in \mathcal{U} : \ y_l \leq z\}$, ir $\sharp A_z$ reiškia A_z elementų skaičių (Krapavickaitė 2002).

Tokia funkcija $F_y(z)$ oficialiojoje statistikoje sutinkama, kai domimasi gyventojų ar namų ūkių pajamų pasiskirstymu. Sudarysime pasiskirstymo funkcijos $F_y(z)$ įvertinį, naudojant dviejų ėmimų schemą, kuri yra pateikta 2.2 paveiksle.

Apibrėžiame indikatorinį kintamąjį $h(z)$ su reikšmėmis

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

$i = 1, 2, \ldots, N$, ir suma $t_{h(z)} = \sum_{i=1}^{N} h_i(z)$.

Tuomet tyrimo kintamojo y populiacijos pasiskirstymo funkcija gali būti užrašyta kaip indikatoriaus $h(z)$ reikšmių $h_i(z)$ vidurkis populiacijoje:

$$F_y(z) = \frac{t_{h(z)}}{N} = \frac{1}{N}\sum_{i=1}^{N} h_i(z). \tag{3.1}$$

Tiriamąją imtį $\mathfrak{s}$ taikant dviejų ėmimų schemą sudaro dviejų imčių atitinkamai s_m ir s_u sąjunga. Be to, kiekviena iš jų yra dviejų fazių imtys.

Naudojant π^* įvertinių konstravimo idėją (Särndal *et al* 1992) tyrimo kintamojo populiacijos sumai vertinti dviejų fazių imties atveju, įvedama tikimybė:

$$\begin{aligned} \pi_i^* &= \mathrm{P}(\mathfrak{s}' : i \in \mathfrak{s}')\mathrm{P}(\mathfrak{s}_m : i \in \mathfrak{s}_m \mid \mathfrak{s}') \\ &\quad + \mathrm{P}(\mathfrak{s}'^c : i \in \mathfrak{s}'^c)\mathrm{P}(\mathfrak{s}_u : i \in \mathfrak{s}_u \mid \mathfrak{s}'^c) \\ &= \pi_i'\pi_{i|\mathfrak{s}'} + \pi_i'^c\pi_{i|\mathfrak{s}'^c}, \end{aligned} \tag{3.2}$$

čia $\pi_i'^c = 1 - \pi_i'$.

Naudojant imčių $\mathfrak{s}_u$ ir $\mathfrak{s}_m$ duomenis galime sudaryti nepaslinktąjį π^* pasisikirstymo funkcijos (3.1) įvertinį

$$\begin{aligned} \widehat{F}_y(z) &= \frac{1}{N}\sum_{i\in s}\frac{h_i(z)}{\pi_i^*} \\ &= \frac{1}{N}\sum_{i\in s_m}\frac{h_i(z)}{\pi_i^*} + \frac{1}{N}\sum_{i\in s_u}\frac{h_i(z)}{\pi_i^*} \\ &= \frac{1}{N}\sum_{i\in s_m}\frac{\pi_i'\pi_{i|s'}}{\pi_i^*}\frac{h_i(z)}{\pi_i'\pi_{i|s'}} + \frac{1}{N}\sum_{i\in s_u}\frac{\pi_i'^c\pi_{i|s'^c}}{\pi_i^*}\frac{h_i(z)}{\pi_i'^c\pi_{i|s'^c}}, \end{aligned} \tag{3.3}$$

kai kiekviename iš ėmimų, renkamos imtys pagal bet kokį imties planą.

Įvedame naujus žymėjimus:

$$d_{1i} = \frac{\pi_i^{'}\pi_{i|\mathfrak{s}'}}{\pi_i^*},\ i \in \mathfrak{s}_m, \quad \hat{t}_{h(z)_m} = \sum_{i\in\mathfrak{s}_m}\frac{h_i(z)}{\pi_i^{'}\pi_{i|\mathfrak{s}'}},$$
$$d_{2i} = \frac{\pi_i^{'c}\pi_{i|\mathfrak{s}'^c}}{\pi_i^*},\ i \in \mathfrak{s}_u, \quad \hat{t}_{h(z)_u} = \sum_{i\in\mathfrak{s}_u}\frac{h_i(z)}{\pi_i^{'c}\pi_{i|\mathfrak{s}'^c}}.$$

$\hat{t}_{h(z)_m}$ ir $\hat{t}_{h(z)_u}$ yra nepaslinktieji indikatorinio kintamojo $h(z)$ sumų įvertiniai. Koeficientai d_{1i}, d_{2i} nepriklauso nuo indekso i kai kiekviename iš ėmimų renkamos lygių tikimybių imtys

$$d_{1i} = d_1,\ i \in s_m, \quad d_{2i} = d_2,\ i \in s_u.$$

Su naujais žymėjimais tyrimo kintamojo y populiacijos pasiskirstymo funkcijos $F_y(z)$ įvertinys (3.3) yra

$$\widehat{F}_y(z) = \frac{1}{N}d_1\hat{t}_{h(z)_m} + \frac{1}{N}d_2\hat{t}_{h(z)_u}. \tag{3.4}$$

Apskačiuotos π reikšmės kai kiekviename iš dviejų ėmimų renkamos paprastosios atsitiktinės imtys pateiktos 2.1.1 skyrelyje. Tuomet atitinkamai koeficientai d_1 ir d_2 apskaičiuojami:

$$d_1 = \frac{m}{n}, \quad d_2 = \frac{u}{n}.$$

Pasiskirstymo funkcijos $F_y(z)$ įvertinys (3.4) gali būti užrašytas

$$\widehat{F}_y(z) = \frac{m}{n}\frac{1}{N}\hat{t}_{h(z)_m} + \frac{u}{n}\frac{1}{N}\hat{t}_{h(z)_u} = \frac{m}{n}\overline{h(z)}_m + \frac{u}{n}\overline{h(z)}_u, \tag{3.5}$$

čia

$$\overline{h(z)}_m = \frac{1}{m}\sum_{i\in s_m} h_i(z), \quad \overline{h(z)}_u = \frac{1}{u}\sum_{i\in s_u} h_i(z).$$

Iš 1.7 teiginio seka, kad pasisikirstymo funkcijos $F_y(z)$ įvertinio $\widehat{F}_y(z)$ dispersija yra

$$\mathrm{D}(\widehat{F}_y(z)) = \left(1 - \frac{n}{N}\right)\frac{s^2_{h(z)}}{n}, \tag{3.6}$$

čia

$$s^2_{h(z)} = \frac{1}{N-1}\sum_{i=1}^{N}(h_i(z) - \mu_{h(z)})^2, \quad \mu_{h(z)} = \frac{1}{N}\sum_{i=1}^{N} h_i(z).$$

Įvertinio $\widehat{F}_y(z)$ dispersijai $\mathrm{D}(\widehat{F}_y(z))$ (3.6) vertinti siūlomas nepasliktasis įvertinys:

$$\widehat{\mathrm{D}}(\widehat{F}_y(z)) = \left(1 - \frac{n}{N}\right)\frac{\hat{s}^2_{h(z)}}{n}, \tag{3.7}$$

čia

$$\hat{s}^2_{h(z)} = \frac{1}{n-1}\sum_{i=1}^{n}\left(h_i(z) - \overline{h(z)}_n\right)^2, \quad \overline{h(z)}_n = \frac{1}{n}\sum_{i=1}^{n} h_i(z).$$

Nežinomas parametras $s^2_{h(z)}$ dispersijos (3.6) išraiškoje keičiamas jo įverčiu $\hat{s}^2_{h(z)}$.

Yra žinoma, kad regresiniame ar santykiniame populiacijos tyrimo kintamojo sumos įvertinyje naudojamas papildomas kintamasis su žinoma populiacijos suma tam tikromis sąlygomis leidžia tiksliau įvertinti tyrimo kintamojo sumą. Taikant dviejų ėmimų schemą, tyrimo kintamojo y pirmojo ėmimo reikšmes žymėsime x_i ir naudosime kaip papildomą informaciją. Sudarysime regresinį, santykinį pasiskirstymo funkcijos $F_y(z)$ įvertinius, siekiant pagerinti nepaslinktąjį pasiskirstymo funkcijos įvertinį $\widehat{F}_y(z)$ (3.3), kuris nenaudoja jokios papildomos informacijos.

3.1.2. Regresinis pasiskirstymo funkcijos įvertinys ir jo apytikslė dispersija

Apibrėžiame naują indikatorinį kintamąjį $g(z)$ su reikšmėmis:

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

$i = 1,2,...,N$, ir suma $t_{g(z)} = \sum_{k=1}^{N} g_i(z)$. Tuomet pasiskirstymo funkcija $F_x(z)$ gali būti užrašyta kaip indikatoriaus $g(z)$ reikšmių $g_i(z)$ vidurkis populiacijoje:

$$F_x(z) = \frac{t_{g(z)}}{N} = \frac{1}{N}\sum_{i=1}^{N} g_i(z). \tag{3.8}$$

Naudojant imčių $\mathfrak{s}'$ ir $\mathfrak{s}_m$ duomenis sudaromas regresinis pasiskirstymo funkcijos $F_y(z)$ įvertinys:

$$\widehat{F}^{reg}_{ym}(z) = \frac{1}{N}\hat{t}^{reg}_{h(z)_m} = \frac{1}{N}\hat{t}_{h(z)_m} + \frac{1}{N}b(\hat{t}_{g(z)_{n'}} - \hat{t}_{g(z)_m}), \qquad (3.9)$$

čia

$$\hat{t}_{h(z)_m} = \sum_{i\in\mathfrak{s}_m} \frac{h_i(z)}{\pi_i\pi_{i|\mathfrak{s}'}}, \quad \hat{t}_{g(z)_m} = \sum_{i\in\mathfrak{s}_m} \frac{g_i(z)}{\pi_i\pi_{i|\mathfrak{s}'}}, \quad \hat{t}_{g(z)_{n'}} = \sum_{i\in s'} \frac{g_i(z)}{\pi_i},$$

b yra koeficientas.

Naudojant antrojo ėmimo imties $\mathfrak{s}_u$ duomenis tyrimo kintamojo y pasiskirstymo funkcijai $F_y(z)$ vertinti taikome nepaslinktąjį įvertinį

$$\widehat{F}_{yu}(z) = \frac{1}{N}\hat{t}_{h(z)_u}, \qquad (3.10)$$

čia

$$\hat{t}_{h(z)_u} = \sum_{i\in\mathfrak{s}_u} \frac{h_i(z)}{\pi'^c_i \pi_{i|\mathfrak{s}'^c}}$$

Paėmę dviejų įvertinių (3.9) ir (3.10) tiesinę kombinaciją gauname sudėtinį regresinį baigtinės populiacijos pasiskirstymo funkcijos $F_y(z)$ įvertinį:

$$\widehat{F}^{reg}_{y}(z) = \omega\frac{1}{N}\hat{t}^{reg}_{h(z)_m} + (1-\omega)\frac{1}{N}\hat{t}_{h(z)_u}, \qquad (3.11)$$

čia ω yra konstanta $(0 < \omega < 1)$.

Įvedame naujus žymėjimus:

$$\mathrm{D}_1 = \mathrm{AD}(\hat{t}^{reg}_{h(z)_m}), \quad \mathrm{D}_2 = \mathrm{D}(\hat{t}_{h(z)_u}), \quad \mathrm{C} = \mathrm{cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u}).$$

Sudėtinio tyrimo kintamojo y pasiskirstymo funkcijos $F_y(z)$ įvertinio (3.11) apytikslė dispersija gali būti išreikšta taip:

$$\mathrm{AD}(\widehat{F}^{reg}_{y}(z)) = \omega^2\frac{1}{N^2}\mathrm{D}_1 + (1-\omega)^2\frac{1}{N^2}\mathrm{D}_2 + 2\omega(1-\omega)\frac{1}{N^2}\mathrm{C}. \quad (3.12)$$

Nario $\hat{t}^{reg}_{h(z)_m}$ (3.11) išraiškoje dispersija $\mathrm{D}(\hat{t}^{reg}_{h(z)_m})$ priklauso nuo pasirenkamo koeficiento b. Kadangi įvertinys $\hat{t}^{reg}_{h(z)_m}$ buvo sudarytas naudojant dviejų fazių ėmimą, optimalus b, kai kiekvienoje iš fazių renkamos paprastosios atsitiktinės imtys, su kuriuo minimizuojama dispersija $\mathrm{D}(\hat{t}^{reg}_{h(z)_m})$ yra

$$b_{opt} = \frac{s_{h(z)g(z)}}{s^2_{g(z)}} = \frac{\sum_{i=1}^{N}(h_i(z) - \mu_{h(z)})(g_i(z) - \mu_{g(z)})}{\sum_{i=1}^{N}(g_i(z) - \mu_{g(z)})^2}, \tag{3.13}$$

čia

$$\mu_{h(z)} = \frac{1}{N}\sum_{i=1}^{N} h_i(z), \quad \mu_{g(z)} = \frac{1}{N}\sum_{i=1}^{N} g_i(z).$$

Kadangi indikatorinių kintamųjų $h(z)$ ir $g(z)$ reikšmės populiacijoje nėra žinomos, negalime apskaičiuoti b_{opt}, todėl jį vertiname naudodami imties $\mathfrak{s}_m$ duomenis

$$\hat{b}_{opt} = \frac{\hat{s}_{h(z)g(z)}}{\hat{s}^2_{g(z)}} = \frac{\sum_{i\in s_m}(h_i(z) - \overline{h(z)}_m)(g_i(z) - \overline{g(z)}_m)}{\sum_{i\in s_m}(g_i(z) - \overline{g(z)}_m)^2}, \tag{3.14}$$

čia

$$\overline{g(z)}_m = \frac{1}{m}\sum_{i\in s_m} g_i(z),$$

$\overline{h(z)}_m$ buvo apibrėžtas (3.5) išraiškoje.

Kai kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys, pasiskirstymo funkcijos $F_y(z)$ sudėtinis regresinis įvertinys (3.11) yra

$$\widehat{F}^{reg}_y(z) = \omega\big(\overline{h(z)}_m + \hat{b}_{opt}(\overline{g(z)}_{n'} - \overline{g(z)}_m)\big) + (1-\omega)\overline{h(z)}_u, \tag{3.15}$$

čia

$$\overline{g(z)}_{n'} = \frac{1}{n'}\sum_{i\in s'} g_i(z), \quad \overline{g(z)}_m = \frac{1}{m}\sum_{i\in s_m} g_i(z),$$

$\overline{h(z)}_m$ ir $\overline{h(z)}_u$ jau buvo apibrėžtos 3.5 lygybėje.

3.1 teiginys. *Jei kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys, pasiskirstymo funkcijos $F_y(z)$ sudėtinio regresinio įvertinio $\widehat{F}^{reg}_y(z)$ (3.15) apytikslė dispersija $\mathrm{AD}(\widehat{F}^{reg}_y(z))$ gali būti išreiškiama taip:*

$$\begin{aligned}\mathrm{AD}(\widehat{F}^{reg}_y(z)) &= \omega^2\frac{1}{N^2}\mathrm{AD}(\hat{t}^{reg}_{h(z)_m}) + (1-\omega)^2\frac{1}{N^2}\mathrm{D}(\hat{t}_{h(z)_u}) \\ &\quad + 2\omega(1-\omega)\frac{1}{N^2}\mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u}),\end{aligned} \tag{3.16}$$

$$\mathrm{AD}(\hat{t}^{reg}_{h(z)_m}) = N^2\left(\left(1-\frac{n'}{N}\right)\frac{s^2_{h(z)}}{n'} + \left(1-\frac{m}{n'}\right)\frac{s^2_{D(z)}}{m}\right),$$

čia

$$\mathrm{D}(\hat{t}_{h(z)_u}) = N^2\left(1-\frac{u}{N}\right)\frac{s^2_{h(z)}}{u}, \quad \mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u}) = -Ns^2_{h(z)},$$

$$s^2_{D(z)} = \frac{1}{N-1}\sum_{i=1}^{N}(D_i(z)-\mu_{D(z)})^2, \quad \mu_{D(z)} = \frac{1}{N}\sum_{i=1}^{N}D_i(z),$$

$s^2_{h(z)}$ *apibrėžtas* (3.7) *lygybėje, o* $D_i(z) = h_i(z) - b_{opt}g_i(z)$.

Įrodymas. Jei kiekviename iš ėmimų imtys renkamos pagal bet kokį imties planą, pasiskirstymo funkcijos $F_y(z)$ sudėtinio regresinio pasiskirstymo funkcijos įvertinio $\widehat{F}^{reg}_y(z)$ (3.15) dispersija yra

$$\begin{aligned}\mathrm{D}(\widehat{F}^{reg}_y(z)) &= \omega^2\frac{1}{N^2}\mathrm{D}(\hat{t}^{reg}_{h(z)_m}) + (1-\omega)^2\frac{1}{N^2}\mathrm{D}(\hat{t}_{h(z)_u}) \\ &+2\omega(1-\omega)\frac{1}{N^2}\mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u}). \end{aligned} \tag{3.17}$$

Apytikslė sumos regresinio įvertinio dispersija dviejų fazių ėmimo atveju, kai kiekvienoje fazėje renkamos imtys, pagal bet kokį imties planą buvo nagrinėta Särndal *et al* (1992). Įvertinio $\hat{t}^{reg}_{h(z)_m}$ apytikslę dispersiją galime išreikšti taip:

$$\begin{aligned}\mathrm{AD}(\hat{t}^{reg}_{h(z)_m}) &= \sum_{i,j\in\mathcal{U}}(\pi'_{ij} - \pi'_i\pi'_j)\frac{h_i(z)}{\pi'_i}\frac{h_j(z)}{\pi'_j} \\ &+\mathrm{E}\left(\sum_{i,j\in s'}(\pi_{ij|s'} - \pi_{i|s'}\pi_{j|s'})\frac{D_i(z)}{\pi'_i\pi_{i|s'}}\frac{D_j(z)}{\pi'_j\pi_{j|s'}}\right),\end{aligned} \tag{3.18}$$

čia $D_i(z) = h_i(z) - b_{opt}g_i(z)$. Atitinkamai

$$\begin{aligned}\mathrm{D}(\hat{t}_{h(z)_u}) &= \sum_{i,j\in\mathcal{U}}(\pi'^c_{ij} - \pi'^c_i\pi'^c_j)\frac{h_i(z)}{\pi'^c_i}\frac{h_j(z)}{\pi'^c_j} \\ &+\mathrm{E}\left(\sum_{i,j\in s'^c}(\pi_{ij|s'^c} - \pi_{i|s'^c}\pi_{j|s'^c})\frac{h_i(z)}{\pi'^c_i\pi_{i|s'^c}}\frac{h_j(z)}{\pi'^c_j\pi_{j|s'^c}}\right),\end{aligned} \tag{3.19}$$

$$\mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u}) = -\sum_{i,j\in\mathcal{U}} (\pi'_{ij} - \pi'_i \pi'_j) \frac{h_i(z)}{\pi'_i} \frac{h_j(z)}{\pi'^c_j}. \tag{3.20}$$

Pakeitę π reikšmes (3.18) - (3.20) išraiškose apskaičiuotomis reikšmėmis, kai kiekviename iš ėmimų renkamos paprastosios atstitiktinės imtys, gauname pasiskirstymo funkcijos įvertinio (3.15) apytikslės dispersijos išraišką (3.16). □

3.1 Pastaba. sudėtinio regresinio pasiskirstymo funkcijos įvertinio (3.15) dispersijai $\mathrm{D}(\widehat{F}^{reg}_y(z))$ vertinti kintamųjų $h(z)$ ir $D(z)$ populiacijos dispersijos atitinkamai $s^2_{h(z)}$ ir $s^2_{D(z)}$ naryje $\mathrm{AD}(\hat{t}^{reg}_{h(z)_m})$ (3.16) keičiamos į

$$\hat{s}^2_{h(z)_m} = \frac{1}{m-1} \sum_{i\in s_m} \left(h_i(z) - \overline{h(z)}_m\right)^2, \quad \overline{h(z)}_m = \frac{1}{m} \sum_{i\in s_m} h_i(z),$$

$$\hat{s}^2_{D(z)_m} = \frac{1}{m-1} \sum_{i\in s_m} \left(\widehat{D}_i(z) - \overline{\widehat{D}(z)}_m\right)^2, \quad \overline{\widehat{D}(z)}_m = \frac{1}{m} \sum_{i\in s_m} \widehat{D}_i(z),$$

čia $\widehat{D}_i(z) = h_i(z) - \widehat{b}_{opt} g_i(z)$.

Naryje $\mathrm{D}(\widehat{t}_{h(z)_u})$, $s^2_{h(z)}$ keičiamas į

$$\widehat{s}^2_{h(z)_u} = \frac{1}{u-1} \sum_{i\in s_u} \left(h_i(z) - \overline{h(z)}_u\right)^2, \quad \overline{h(z)}_u = \frac{1}{u} \sum_{i\in s_u} h_i(z),$$

ir naryje $\mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u})$, $s^2_{h(z)}$ atitinkamai keičiamas į

$$\hat{s}^2_{h(z)_n} = \frac{1}{n-1} \sum_{i\in s} \left(h_i(z) - \overline{h(z)}_n\right)^2, \quad \overline{h(z)}_n = \frac{1}{n} \sum_{i\in s} h_i(z).$$

Sudėtinio regresinio pasiskirstymo funkcijos įvertinio $\widehat{F}^{reg}_y(z)$ (3.11) išraiškoje naudojama laisvai pasirenkama konstanta ω, $0 < \omega < 1$. Optimali jos reikšmė ω_{opt} gali būti surasta minimizuojant dispersiją (3.12).

3.2 teiginys. *Naudojant dviejų ėmimų schemą, kai kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys, konstanta ω_{opt} apskaičiuojama sekančiai:*

$$\omega_{opt} = \frac{\mathrm{D}(\hat{t}_{h(z)_u}) - \mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u})}{\mathrm{AD}(\hat{t}^{reg}_{h(z)_m}) + \mathrm{D}(\hat{t}_{h(z)_u}) - 2\mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u})} \tag{3.21}$$

Įrodymas yra panašus kaip ir 2.1 teiginio.

Pakeitę koeficientą ω koeficientu ω_{opt} sudėtinio regresinio pasiskirstymo funkcijos įvertinio $\widehat{F}_y^{reg}(z)$ išraiškoje (3.11), gauname optimalų pasiskirstymo funkcijos įvertinį

$$\widehat{F}_{y\ opt}^{reg}(z) = \omega_{opt}(\overline{h(z)}_m + \hat{b}_{opt}(\overline{g(z)}_{n'} - \overline{g(z)}_m) + (1-\omega_{opt})\overline{h(z)}_u. \quad (3.22)$$

3.3 teiginys. *Paprastosios atsitiktinės imties atveju regresinio pasiskirstymo funkcijos* $F_y(z)$ *įvertinio* $\widehat{F}_{y\ opt}^{reg}(z)$ *apytikslės dispersijos* $\mathrm{AD}(\widehat{F}_{y\ opt}^{reg}(z))_{min}$ *išraiška yra*

$$\mathrm{AD}(\widehat{F}_{y\ opt}^{reg}(z))_{min} = \frac{1}{N^2}\left(\frac{\mathrm{D}_1\mathrm{D}_2 - \mathrm{Cov}^2}{\mathrm{D}_1 + \mathrm{D}_2 - 2\mathrm{Cov}}\right). \quad (3.23)$$

Čia $\mathrm{D}_1 = \mathrm{AD}(\hat{t}_{h(z)_m}^{reg})$, $\mathrm{D}_2 = \mathrm{D}(\hat{t}_{h(z)_u})$, $\mathrm{Cov} = \mathrm{Cov}(\hat{t}_{h(z)_m}^{reg}, \hat{t}_{h(z)_u})$.

Įrodymas. Pakeitę koeficientą ω surasta optimalia išraiška ω_{opt} (3.21) apytikslės dispersijos išraiškoje (3.12) gauname (3.23). □

3.2 Pastaba. Koeficientas w_{opt} vertinamas, pakeičiant tikrąsias dispersijas ir kovariaciją atitinakamais jų įvertiniais

$$\widehat{\omega}_{opt} = \frac{\widehat{\mathrm{D}}(\hat{t}_{h(z)_u}) - \widehat{\mathrm{Cov}}(\hat{t}_{h(z)_m}^{reg}, \hat{t}_{h(z)_u})}{\widehat{\mathrm{D}}(\hat{t}_{h(z)_m}^{reg}) + \widehat{\mathrm{D}}(\hat{t}_{h(z)_u}) - 2\widehat{\mathrm{Cov}}(\hat{t}_{h(z)_m}^{reg}, \hat{t}_{h(z)_u})}. \quad (3.24)$$

Sudėtinio regresinio įvertinio (3.22) dispersija $\mathrm{AD}(\widehat{F}_{y\ opt}^{reg}(z))_{min}$ vertinama narius D_1, D_2, and Cov išraiškoje (3.23) keičiant atitinkamais įvertiniais $\widehat{\mathrm{D}}_1$, $\widehat{\mathrm{D}}_2$ ir $\widehat{\mathrm{Cov}}$.

3.1.3. Santykinis pasiskirstymo funkcijos įvertinys ir jo apytikslė dispersija

Atskiras regresinio įvertinio atvejis yra santykinis įvertinys. Skirtumas tarp tyrimo kintamojo y populiacijos pasiskirstymo funkcijos regresinio įvertinio ir santykinio įvertinio yra koeficiento b (3.9) parinkimas.

Naudojant imčių $\mathfrak{s}'$ ir $\mathfrak{s}_m$ duomenis sudaromas santykinis pasiskirstymo funkcijos $F_y(z)$ įvertinys:

$$\widehat{F}_{ym}^{sant}(z) = \frac{1}{N}\hat{t}_{h(z)_m}^{sant} = \frac{1}{N}\hat{t}_{g(z)_{n'}}\widehat{R}(z), \quad (3.25)$$

čia

$$\hat{t}_{g(z)_{n'}} = \sum_{i\in\mathfrak{s}'} \frac{g_i(z)}{\pi_i'}, \quad \widehat{R}(z) = \frac{\hat{t}_{h(z)_m}}{\hat{t}_{g(z)_m}},$$
$$\hat{t}_{h(z)_m} = \sum_{i\in\mathfrak{s}_m} \frac{h_i(z)}{\pi_i'\pi_{i|\mathfrak{s}'}}, \quad \hat{t}_{g(z)_m} = \sum_{i\in\mathfrak{s}_m} \frac{g_i(z)}{\pi_i'\pi_{i|\mathfrak{s}'}},$$

kur b parenkamas atitinkamai $b = \frac{\hat{t}_{h(z)_m}}{\hat{t}_{g(z)_m}} = \widehat{R}(z)$.

Antrasis pasiskirstymo funkcijos $F_y(z)$ įvertinys $\widehat{F}_{yu}(z)$ (3.10) apibrėžiamas naudojant imties s_u duomenis. Paėmę dviejų įvertinių (3.25) ir (3.10) tiesinę kombinaciją sudarome sudėtinį santykinį baigtinės populiacijos pasiskirstymo funkcijos $F_y(z)$ įvertinį

$$\widehat{F}_y^{sant}(z) = \lambda\frac{1}{N}\hat{t}^{sant}_{h(z)_m} + (1-\lambda)\frac{1}{N}\hat{t}_{h(z)_u}, \tag{3.26}$$

čia λ yra konstanta $(0 < \lambda < 1)$.

Tarkime, kad kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys. Tuomet pasiskirstymo funkcijos $F_y(z)$ sudėtinis santykinis įvertinys (3.26) gali būti užrašytas

$$\widehat{F}_y^{sant}(z) = \lambda\overline{g(z)}_{n'}\widehat{R}(z) + (1-\lambda)\overline{h(z)}_u, \tag{3.27}$$

čia

$$\widehat{R}(z) = \frac{\sum_{i\in\mathfrak{s}_m} h_i(z)}{\sum_{i\in\mathfrak{s}_m} g_i(z)}.$$

3.4 teiginys. *Jei kiekviename iš dviejų ėmimų renkamos paprastosios atsitiktinės imtys, pasiskirstymo funkcijos $F_y(z)$ sudėtinio santykinio įvertinio (3.27) apytikslė dispersija $\mathrm{AD}(\widehat{F}_y^{sant}(z))$ turi pavidalą*

$$\begin{aligned}\mathrm{AD}(\widehat{F}_y^{sant}(z)) &= \lambda^2\frac{1}{N^2}\mathrm{AD}(\hat{t}^{sant}_{h(z)_m}) + (1-\lambda)^2\frac{1}{N^2}\mathrm{D}(\hat{t}_{h(z)_u}) \\ &\quad +2\lambda(1-\lambda)\frac{1}{N^2}\mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u}),\end{aligned} \tag{3.28}$$

čia

$$\mathrm{AD}(\hat{t}^{sant}_{h(z)_m}) = N^2\left(\left(1-\frac{n'}{N}\right)\frac{s^2_{h(z)}}{n'} + \left(1-\frac{m}{n'}\right)\frac{s^2_{R(z)}}{m}\right), \tag{3.29}$$

$$s^2_{R(z)} = \frac{1}{N-1}\sum_{i=1}^{N}(h_i(z) - R(z)g_i(z))^2, \quad R(z) = \frac{\sum_{i=1}^{N} h_i(z)}{\sum_{i=1}^{N} g_i(z)},$$
$$\mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u}) = -Ns^2_{h(z)},$$

$\mathrm{D}(\hat{t}_{h(z)_u})$ *ir* $s^2_{h(z)}$ *yra pateiktos* (3.16) *ir* (3.6) .

Įrodymas. Sudėtinio santykinio pasiskirstymo funkcijos $F_y(z)$ įvertinio (3.26) dispersija užrašoma

$$\begin{aligned} \mathrm{D}(\widehat{F}^{sant}_y(z)) &= \lambda^2\frac{1}{N^2}\mathrm{D}(\hat{t}^{sant}_{h(z)_m}) + (1-\lambda)^2\frac{1}{N^2}\mathrm{D}(\hat{t}_{h(z)_u}) \\ &\quad +2\lambda(1-\lambda)\frac{1}{N^2}\mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u}). \end{aligned} \tag{3.30}$$

Apytikslė $\mathrm{D}(\hat{t}^{sant}_{h(z)_m})$ dispersijos išraiška gaunama skleidžiant įvertinį $\hat{t}^{sant}_{h(z)_m}$ Teiloro eilute ir imant tiesinę šio skleidinio dalį:

$$\begin{aligned} \mathrm{AD}(\hat{t}^{sant}_{h(z)_m}) &= \sum_{i,j\in\mathcal{U}}(\pi'_{ij} - \pi'_i\pi'_j)\frac{h_i(z)}{\pi'_i}\frac{h_j(z)}{\pi'_j} \\ &\quad +\mathrm{E}\left(\sum_{i,j\in\mathfrak{s}'}(\pi_{ij|\mathfrak{s}'} - \pi_{i|s'}\pi_{j|\mathfrak{s}'})\frac{R_i(z)}{\pi'_i\pi_{i|\mathfrak{s}'}}\frac{R_j(z)}{\pi'_j\pi_{j|\mathfrak{s}'}}\right), \end{aligned} \tag{3.31}$$

čia $R_i(z) = h_i(z) - R(z)g_i(z)$. $\mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u}) = \mathrm{Cov}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u})$ ir $\mathrm{D}(\hat{t}_{h(z)_u})$ išraiškos pateiktos atitinkamai (3.20) ir (3.19). $R(z)$ apibrėžta (3.28) lygybe.

Pakeisdami π reikšmes atitinkamomis, naudojant paprastosios atsitiktinės imties planą kiekviename iš dviejų ėmimų, gauname pasiskirstymo funkcijos $F_y(z)$ sudėtinio santykinio įvertinio (3.27) apytikslės dispersijos išraišką (3.4) □

3.3 Pastaba. Pasiskirstymo funkcijos $F_y(z)$ sudėtinio santykinio pasiskirstymo funkcijos įvertinio (3.27) dispersija vertinama išraiškoje $\mathrm{AD}(\hat{t}^{sant}_{h(z)_m})$ (3.28) pakeičiant $s^2_{h(z)}$, $s^2_{R(z)}$ atitinkamais įverčiais

$$\hat{s}^2_{R(z)_m} = \frac{1}{m-1}\sum_{i\in s_m}\left(h_i(z) - \widehat{R}(z)g_i(z)\right)^2, \quad \widehat{R}(z) = \frac{\sum_{i\in s_m} h_i(z)}{\sum_{i\in s_m} g_i(z)}.$$

Atitinkamai įvertiniai $\widehat{\mathrm{D}}(\hat{t}_{h(z)_u})$ ir $\widehat{\mathrm{Cov}}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u}) = \widehat{\mathrm{Cov}}(\hat{t}^{reg}_{h(z)_m}, \hat{t}_{h(z)_u})$ buvo pasiūlyti nagrinėjant sudėtinio regresinio pasiskirstymo funkcijos dispersijos įvertinį $\widehat{D}(\widehat{F}^{reg}_y(z))$ (3.1 pastaba).

Jei kiekviename iš ėmimų renkamos paprastosios atsitiktinės imtys, optimalus sudėtinis santykinis pasiskirstymo funkcijos įvertinys yra

$$\widehat{F}^{sant}_{y\ opt}(z) = \lambda_{opt}\overline{g(z)}_{n'}\widehat{R}(z) + (1-\lambda_{opt})\frac{1}{N}\overline{h(z)}_u, \tag{3.32}$$

čia

$$\lambda_{opt} = \frac{\mathrm{D}(\hat{t}_{h(z)_u}) - \mathrm{Cov}(\hat{t}^r_{h(z)_m}, \hat{t}_{h(z)_u})}{\mathrm{AD}(\hat{t}^{sant}_{h(z)_m}) + \mathrm{Var}(\hat{t}_{h(z)_u}) - 2\mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u})}.$$

Apytikslė minimali santykinio pasiskirstymo funkcijos įvertinio $\widehat{F}^{sant}_{y\ opt}(z)$ dispersija $\mathrm{AD}(\widehat{F}^{r}_{y\ opt}(z))_{min}$ apibrėžiama taip:

$$\mathrm{AD}(\widehat{F}^{sant}_{y\ opt}(z))_{min} = \frac{1}{N^2}\left(\frac{\mathrm{D}_1\mathrm{D}_2 - \mathrm{Cov}^2}{\mathrm{D}_1 + \mathrm{D}_2 - 2\mathrm{Cov}}\right), \tag{3.33}$$

čia $\mathrm{D}_1 = \mathrm{AD}(\hat{t}^{sant}_{h(z)_m})$, $\mathrm{D}_2 = \mathrm{D}(\hat{t}_{h(z)_u})$, $\mathrm{Cov} = \mathrm{Cov}(\hat{t}^{sant}_{h(z)_m}, \hat{t}_{h(z)_u})$.

3.4 Pastaba. Koeficientą λ_{opt} ir pasisikirstymo funkcijos įvertinio $\widehat{F}^{sant}_{y\ opt}(z))$ dispersiją $\mathrm{AD}(\widehat{F}^{sat}_{y\ opt}(z))_{min}$ vertiname, pakeičiant tikrąsias dispersijas ir kovariaciją atitinkamais įvertiniais.

3.1.4. Įvertinių tikslumo tyrimas atliekant matematinį modeliavimą

Toliau atliksime matematinį modeliavimą, kurio metu palyginsime įvertinių, sudarytų naudojant dviejų ėmimų schemą, kai kiekviename iš ėmimų renkama paprastoji atsitiktinė imtis, tikslumą.

Modeliavimui naudojami Statistikos departamento 2005 ir 2006 metų pajamų ir gyvenimo sąlygų statistinio tyrimo duomenys. Tyrimo kintamasis y – namų ūkio bendrosios pajamos. Nagrinėjama N=2 932 namų ūkių populiacija, kiekvienas iš namų ūkių yra indentifikuotas. Duomenys tenkina dviejų ėmimų schemą. Tyrimo kintamojo y reikšmes x_i pirmajame ėmime atitiks 2005 metų tyrimo kintamojo reikšmės, atitinkami tyrimo kintamojo y reikšmes y_i antrajame ėmime atitiks 2006 metų tyrimo kintamojo reikšmės. Koreliacijos koeficientas tarp tyrimo kintamojo y reikšmių 2005 ir 2006 metų yra $\varrho(x, y)$=0,86. Turima pakankamai stipri priklausomybė. Sudarant įvertinį $\widehat{F}_y(z)$ parenkame konkrečius taškus z_k:

$$z_1 = K_{0,10}, \quad z_2 = K_{0,25}, \quad z_3 = K_{0,50}, \quad z_4 = K_{0,75}, \quad z_5 = K_{0,90},$$

čia K_q yra tyrimo kintamojo y nagrinėjamoje populiacijoje q lygio kvantilis.

Renkame B=10 000 $n^{'} = 200$ dydžio paprastųjų atsitiktinių imčių pirmajame ėmime, su skirtingo dydžio paprastosiomis atsitiktinėmis imtimis antrajame ėmime: $\frac{m}{n} = \frac{1}{4}$ (m=50, u=150), $\frac{m}{n} = \frac{1}{2}$ (m=100, u=100) ir $\frac{m}{n} = \frac{3}{4}$ (m=150, u=50). Skaičiuojant pasiskirstymo funkcijos įverčius naudojami pasiūlyti įvertiniai: tradicinis įvertinys $\widehat{F}_y(z)$, santykinis ir regresinis įvertiniai $\widehat{F}_y^{sant}(z)$ ir $\widehat{F}_y^{reg}(z)$ atitinkamai, su koeficientais $\omega = 0{,}5$ ir $\lambda = 0{,}5$, taip pat optimalus santykinis ir regresinis įvertiniai $\widehat{F}_{y\ opt}^{sant}(z)$ ir $\widehat{F}_{y\ opt}^{reg}(z)$ atitinkamai, turintys mažiausias dispersijas su optimaliais koeficientais ω_{opt} ir λ_{opt}.

Naudojant kiekvieną iš įvertinių skaičiuojami tyrimo kintamojo y pasiskirstymo funkcijos įverčiai taškuose $K_{0.10}$, $K_{0,25}$, $K_{0,50}$, $K_{0,75}$, ir $K_{0,90}$ apskaičiuotuose iš populiacijos duomenų. Vadinasi, taikant kiekvieną įvertinį gauname 10 000 įverčių. Naudojant šiuos įverčius skaičiuojame santykinius poslinkius, santykines vidutines kvadratines paklaidas, ir santykinius efektyvumus konstruotiems įvertiniams. Kiekvienam įvertiniui $\widehat{\theta}_y(z)$ santykinis poslinkis apibrėžiamas taip

$$\mathrm{SPosl}(\widehat{\theta}_y(z)) = \frac{1}{F_y(z)}\left(\frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_y^{(i)}(z) - F_y(z))\right),$$

santykinė vidutinė kvadratinė paklaida

$$\mathrm{SVKP}(\widehat{\theta}_y(z)) = \frac{1}{F_y(z)}\sqrt{\frac{1}{B}\sum_{i=1}^{B}(\widehat{\theta}_y^{(i)}(z) - F_y(z))^2}.$$

Čia $\widehat{\theta}_y^{(i)}(z)$ yra i - asis įvertis taške z, apskaičiuotas naudojant įvertinį $\widehat{\theta}_y(z)$. Santykinis efektyvumas apibrėžiamas

$$\mathrm{SE}(\widehat{\theta}_y(z)) = \frac{\mathrm{SVKP}(\widehat{F}_y(z))}{\mathrm{SVKP}(\widehat{\theta}_y(z))}.$$

Čia $\mathrm{SVKP}(\widehat{F}_y(z))$ yra tradicinio pasiskirstymo funkcijos $F_y(z)$ įvertinio $\widehat{F}_y(z)$

“VikChaThesis” — 2009/12/21 — 13:05 — page 74 — #90

santykinė vidutinė kvadratinė paklaida. Taip pat kiekvienam įvertiniui apibrėžiamas efektyvumas $\mathrm{E}(\widehat{\theta}_y(z))$, kuris yra nagrinėjamų pasiskirstymo funkcijos įvertinių dispersijos įvertinių santykis $\widehat{\mathrm{D}}(\widehat{F}_y(z))$ ir $\widehat{\mathrm{D}}(\widehat{\theta}_y(z))$.

3.1 lentelėje pateikti nagrinėtų populiacijos pasiskirstymo funkcijos įvertinių santykiniai poslinkiai.

3.1 lentelė. Reali populiacija: įvertinių santykinis poslinkis (SPosl)

Table 3.1. Real population: relative bias of estimators

Įvertinys $\hat{\theta}$	$K_{0,10}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,90}$
$\frac{m}{n} = \frac{1}{4}$					
$\hat{F}_y$	0,01148	0,00009	−0,00052	−0,00038	−0,00041
$\hat{F}_y^{sant}$	0,02512	0,00394	0,00086	0,00019	−0,00010
$\hat{F}_y^{reg}$	0,01131	0,00077	−0,00009	−0,00013	−0,00046
$\hat{F}_{y\ opt}^{sant}$	−0,01278	−0,00287	0,00148	0,00348	0,00376
$\hat{F}_{y\ opt}^{reg}$	−0,02036	−0,00478	0,00117	0,00361	0,00403
$\frac{m}{n} = \frac{1}{2}$					
$\hat{F}_y$	0,01429	−0,00055	−0,00025	0,00015	0,00015
$\hat{F}_y^{sant}$	0,01614	0,00171	0,00063	0,00031	0,00014
$\hat{F}_y^{reg}$	0,01504	0,00046	0,00027	0,00022	0,00012
$\hat{F}_{y\ opt}^{sant}$	−0,01740	−0,00546	0,00082	0,00300	0,00379
$\hat{F}_{y\ opt}^{reg}$	−0,01804	−0,00689	0,00049	0,00298	0,00393
$\frac{m}{n} = \frac{1}{4}$					
$\hat{F}_y$	0,01636	0,00112	0,00135	0,00025	0,00045
$\hat{F}_y^{sant}$	0,01682	0,00153	0,00146	0,00023	0,00040
$\hat{F}_y^{reg}$	0,01560	0,00129	0,00141	0,00024	0,00040
$\hat{F}_{y\ opt}^{sant}$	−0,02234	−0,00827	0,00105	0,00313	0,00500
$\hat{F}_{y\ opt}^{reg}$	−0,02415	−0,00865	0,00098	0,00315	0,00502

Neoptimaliems pasiskirstymo funkcijos įvertiniams santykinis poslinkis mažėja didėjant baigtinės populiacijos kvantilio lygiui q. Tradicinio ir sudėtinio regresinio pasiskirstymo funkcijos įvertinių, atitinkamai $\widehat{F}_y(z)$ ir $\widehat{F}_y^{reg}(z)$, rezultatai panašūs, bet įvertinys $\widehat{F}_y^{reg}(z)$ turi mažesnį santykinį poslinkį. Sudėtinis regresinis pasiskirstymo funkcijos įvertinys $\widehat{F}_y^{reg}(z)$ turi mažesnį poslinkį lyginant su sudėtiniu santykiniu pasiskirstymo funkcijos įvertiniu $\widehat{F}_y^{sant}(z)$, ypatingai su mažais baigtinės populiacijos q lygio kvantiliais, o taip pat su mažesniu santykiu m/n. Optimalūs sudėtiniai įvertiniai regresinis ir santykinis, atitinkamai $\widehat{F}_{y\ opt}^{sant}(z)$ ir $\widehat{F}_{y\ opt}^{reg}(z)$ visais atvejais turi didžiausius santykinius poslinkius.

3.2 lentelėje pateiktos įvertinių santykinės vidutinės kvadratinės paklaidos.

3.2 lentelė. Reali populiacija: įvertinių santykinė vidutinė kvadratinė paklaida (SVKP)

Table 3.2. Real population: relative root mean square error of estimators

Įvertinys $\hat{\theta}$	$K_{0,10}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,90}$
$\frac{m}{n} = \frac{1}{4}$					
$\widehat{F}_y$	0,2066	0,1206	0,0688	0,0397	0,0233
$\widehat{F}_y^{sant}$	0,2340	0,1192	0,0667	0,0393	0,0233
$\widehat{F}_y^{reg}$	0,2160	0,1171	0,0659	0,0392	0,0299
$\widehat{F}_{y\ opt}^{sant}$	0,2207	0,1136	0,0639	0,0378	0,0299
$\widehat{F}_{y\ opt}^{reg}$	0,2173	0,1131	0,0637	0,0381	0,0233
$\frac{m}{n} = \frac{1}{2}$					
$\widehat{F}_y$	0,2083	0,1195	0,0693	0,0401	0,0299
$\widehat{F}_y^{sant}$	0,1996	0,1115	0,0642	0,0376	0,0214
$\widehat{F}_y^{reg}$	0,1969	0,1110	0,0640	0,0375	0,0214
$\widehat{F}_{y\ opt}^{sant}$	0,2065	0,1119	0,0638	0,0378	0,0223
$\widehat{F}_{y\ opt}^{reg}$	0,2040	0,1113	0,0636	0,0378	0,0224
$\frac{m}{n} = \frac{3}{4}$					
$\widehat{F}_y$	0,2069	0,1194	0,0687	0,0394	0,0228
$\widehat{F}_y^{sant}$	0,2377	0,1371	0,0787	0,0455	0,0264
$\widehat{F}_y^{reg}$	0,2368	0,1370	0,0786	0,0454	0,0263
$\widehat{F}_{y\ opt}^{sant}$	0,2192	0,1159	0,0655	0,0385	0,0244
$\widehat{F}_{y\ opt}^{reg}$	0,2175	0,1155	0,0654	0,0385	0,0244

Lygindami įvertinių efektyvumą pagal šį tikslumo matą pastebime, kad sudėtinis regresinis pasiskirstymo funkcijos įvertinys $\widehat{F}_y^{reg}(z)$ yra efektyvesnis lyginant su sudėtiniu santykiniu pasiskirstymo funkcijos įvertiniu $\widehat{F}_y^{sant}(z)$, ypatingai mažiems q lygio baigtinės populiacijos kvantiliams ir mažam santykiui m/n. Tai tikriausiai įtakoja specifinės indikatorinių kintamųjų, atitinkamai $g(z)$ ir $h(z)$, reikšmės. Daugeliu atveju tradicinis pasiskirstymo funkcijos įvertinys $\widehat{F}_y(z)$, turi didžiausią santykinę vidutinę kvadratinę paklaidą ypatingai su mažais kvantilių lygiais q. Optimalūs įvertiniai atitinkamai $\widehat{F}_{y\ opt}^{sant}(z)$ ir $\widehat{F}_{y\ opt}^{reg}(z)$ daugeliu atveju yra efektyvesni, turi mažesnę santykinę vidutinę kvadratinę paklaidą, negu atitinkami $\widehat{F}_y^{sant}(z)$ ir $\widehat{F}^{reg}(z)$, įvertiniai.

Santykinis efektyvumas vertintas kiekvienam įvertiniui $\widehat{F}_y^{sant}(z)$, $\widehat{F}_y^{reg}(z)$, $\widehat{F}_{y\ opt}^{sant}(z)$, $\widehat{F}_{y\ opt}^{reg}(z)$ ir standartiniam pasiskirstymo funkcijos įvertiniui $\widehat{F}_y(z)$

pateiktas 3.3 lentelėje.

3.3 lentelė. Reali populiacija: įvertinių santykinis efektyvumas (SE)

Table 3.3. Real population: relative efficiency of estimators

Įvertinys $\hat{\theta}$	$K_{0,10}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,90}$
$\frac{m}{n}=\frac{1}{4}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	0,883	1,012	1,032	1,012	1,003
$\widehat{F}_y^{reg}$	0,956	1,030	1,043	1,014	1,017
$\widehat{F}_{y\ opt}^{sant}$	0,936	1,062	1,076	1,050	1,020
$\widehat{F}_{y\ opt}^{reg}$	0,951	1,067	1,079	1,044	0,999
$\frac{m}{n}=\frac{1}{2}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	1,044	1,072	1,080	1,067	1,070
$\widehat{F}_y^{reg}$	1,058	1,077	1,083	1,069	1,070
$\widehat{F}_{y\ opt}^{sant}$	1,009	1,068	1,087	1,060	1,027
$\widehat{F}_{y\ opt}^{reg}$	1,021	1,074	1,090	1,062	1,021
$\frac{m}{n}=\frac{3}{4}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	0,871	0,871	0,873	0,867	0,866
$\widehat{F}_y^{reg}$	0,874	0,872	0,874	0,868	0,867
$\widehat{F}_{y\ opt}^{sant}$	0,944	1,031	1,048	1,023	0,936
$\widehat{F}_{y\ opt}^{reg}$	0,952	1,034	1,051	1,025	0,937

Optimalūs pasiskirstymo funkcijos įvertiniai $\widehat{F}_{y\ opt}^{sant}(z)$ ir $\widehat{F}_{y\ opt}^{reg}(z)$ yra efektyvesni už atitinkamus $\widehat{F}_y^{sant}(z)$ ir $\widehat{F}_y^{reg}(z)$ įvertinius daugeliu atvejų. Optimalūs įvertiniai turi didesnius santykinius poslinkius, lyginant su kitais pasiskirstymo funkcijos įvertiniais, tai gali būti efektyvumo sumažėjimo priežastis. Tradicinio pasiskirstymo funkcijos įvertinio santykinis efektyvumas yra didžiausias mažiems q lygio populiacijos kvantiliams su mažu ir dideliu santykiu m/n. Optimalių pasiskirstymo funkcijos įvertinių santykinis efektyvumas yra didžiausias su bet kokiu santykiu m/n medianai.

Efektyvumas, kurį parodo tradicinio pasiskirstymo funkcijos $F_y(z)$ dispersijos įvertinio $\widehat{\mathrm{D}}(\widehat{F}_y(z))$ ir atitinkamų dispersijos įvertinių $\widehat{\mathrm{D}}(\widehat{F}_y^{sant}(z))$, $\widehat{\mathrm{D}}(\widehat{F}_y^{reg}(z))$, $\widehat{\mathrm{D}}(\widehat{F}_{y\ opt}^{sant}(z))$, $\widehat{\mathrm{D}}(\widehat{F}_{y\ opt}^{reg}(z))$ santykis, pateiktas 3.4 lentelėje. Pasiūlytų optimalių pasiskirstymo funkcijų įvertinių efektyvumas (E) didėja lyginant su santykiniu efektyvumu (SE). Pasiūlytų optimalių sudėtinių pasiskirstymo

3.4 lentelė. Reali populiacija: Įvertinių efektyvumas (E)

Table 3.4. Real population: efficiency of estimators

Įvertinys $\hat{\theta}$	$K_{0,10}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,90}$
$\frac{m}{n} = \frac{1}{4}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	0,781	1,004	1,075	1,017	1,002
$\widehat{F}_y^{reg}$	0,995	1,074	1,113	1,043	1,039
$\widehat{F}_{y\ opt}^{sant}$	1,133	1,163	1,195	1,171	1,219
$\widehat{F}_{y\ opt}^{reg}$	1,187	1,196	1,213	1,183	1,237
$\frac{m}{n} = \frac{1}{2}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	1,092	1,148	1,170	1,145	1,139
$\widehat{F}_y^{reg}$	1,126	1,163	1,179	1,151	1,150
$\widehat{F}_{y\ opt}^{sant}$	1,141	1,180	1,204	1,177	1,206
$\widehat{F}_{y\ opt}^{reg}$	1,183	1,200	1,216	1,185	1,221
$\frac{m}{n} = \frac{3}{4}$					
$\widehat{F}_y$	1,000	1,000	1,000	1,000	1,000
$\widehat{F}_y^{sant}$	0,755	0,761	0,764	0,761	0,760
$\widehat{F}_y^{reg}$	0,758	0,762	0,765	0,762	0,761
$\widehat{F}_{y\ opt}^{sant}$	1,119	1,121	1,129	1,117	1,152
$\widehat{F}_{y\ opt}^{reg}$	1,135	1,128	1,134	1,121	1,159

funkcijos įvertinių dispersijos įverčių vidurkis yra mažesnis už empirinių dispersijų vidurkį. Teiloro ištiesinimo metodas naudojamas apytikslėms šių įvertinių dispersijoms rasti. Jeigu skleidžiant nagrinėjamus įvertinius Teiloro eilute ieškant apytikslių dispersijų išraiškas imtume aukštesnių eilių narius negu pirmosios eilės tikriausiai šių įvertinių apytiksles dispersijas gautume tikslesnes.

3.2. Tiesioginiai kvantilių įvertiniai

Nagrinėkime baigtinę populiaciją $\mathcal{U} = \{1,2,\ldots,N\}$ su tyrimo kintamuoju y, kurio reikšmės $\{y_1, y_2, \ldots, y_N\}$. Baigtinėje populiacijoje apibrėžto tyrimo kintamojo y q lygmens kvantilis apibrėžiamas

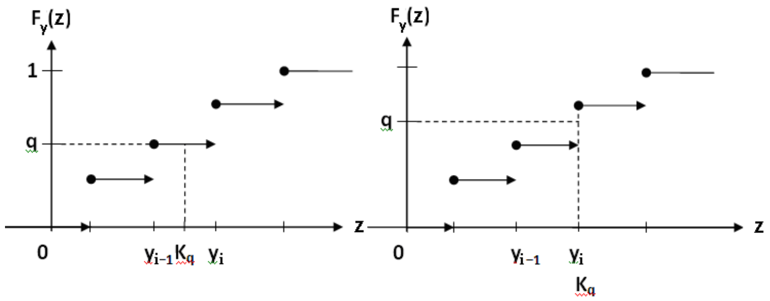
$$K_{yq} = \inf\{z : F_y(z) \geq q\}. \tag{3.34}$$

Baigtinės populiacijos tyrimo kintamojo y skirstinio kvantilį galima apibrėžti ir kitaip. Tyrimo kintamojo y reikšmes išrikiuojame didėjimo tvarka $y_1 \leq y_2 \leq ... \leq y_N$. Baigtinės populiacijos q-lygio kvantilį K_{yq}, $0 < q < 1$, apibrėžkime

$$K_{yq} = F_y^{-1}(q) = \begin{cases} 0{,}5(y_{i-1} + y_i), & \text{jeigu } F(y_i) = q; \\ y_i, & \text{jeigu } F(y_{i-1}) < q < F(y_i), \end{cases} \tag{3.35}$$

su kiekvienu $i \in \mathcal{U}$.

Lygties (3.35) sprendiniai parodyti 3.1 paveiksle.



(a) (b)

3.1 pav. Populiacijos kvantilio nustatymas
Fig. 3.1. Determination of the population quantile

Nagrinėkime n dydžio paprastąją atsitiktinę imtį $\mathfrak{s}$ išrinktą iš baigtinės populiacijos $\mathcal{U}$. Tyrimo kintamojo y reikšmes imtyje išrikiuojame didėjimo tvarka $y_1 \leq y_2 \leq ... \leq y_n$.

Tyrimo kintamojo pasiskirstymo funkcijos $F_y(z)$ įvertinys paprastosios atsitiktinės imties atveju yra

$$\widehat{F}_y(z) = \frac{\hat{t}_y(z)}{N}, \tag{3.36}$$

čia

$$\hat{t}_y(z) = \frac{N}{n} \sum_{i \in \mathfrak{s}} y_i(z),$$

$$y_i(z) = \begin{cases} 1, & \text{jeigu } y_i \leq z, \\ 0, & \text{jeigu } y_i > z. \end{cases}$$

Baigtinės populiacijos kvantilio įvertinį naudodami pasiskirstymo funkcijos $F_y(z)$ įvertinį $\widehat{F}_y(z)$ (3.36) apibrėžiame taip

$$\widehat{K}_{yq} = \widehat{F}_y^{-1}(q) = \begin{cases} 0{,}5(y_{i-1} + y_i), & \text{jeigu } \widehat{F}(y_i) = q; \\ y_{i+1}, & \text{jeigu } \widehat{F}(y_{i-1}) < q < \widehat{F}(y_i), \end{cases} \tag{3.37}$$

su kiekvienu $i \in \mathfrak{s}$.

Ankstesniuose skyreliuose pasiūlėme sudėtinį santykinį ir sudėtinį regresinį pasiskirstymo funkcijos įvertinius atitinkamai $\widehat{F}_y^{sant}(z)$ (3.27) ir $\widehat{F}_y^{reg}(z)$ (3.15). Tyrimo kintamojo y q lygio kvantilis naudojant dviejų ėmimų schemą gali būti įvertintas tiesiogiai:

$$\widehat{K}_{yq}^{sant} = \widehat{F}_y^{sant\ -1}(q), \quad \widehat{K}_{yq}^{reg} = \widehat{F}_y^{reg\ -1}(q).$$

3.2.1. Įvertinių empirinis palyginimas

Modeliavimui naudojami Statistikos departamento 2005 ir 2006 metų pajamų ir gyvenimo sąlygų statisitinio tyrimo duomenys. Tyrimo kintamasis y – namų ūkio bendrosios pajamos.

Renkame B=1 000 $n^{'} = 200$ dydžio paprastųjų atsitiktinių imčių pirmajame ėmime, su skirtingo dydžio paprastosiomis atsitiktinėmis imtimis antrajame ėmime: $\frac{m}{n} = \frac{1}{4}$ (m=50, u=150), $\frac{m}{n} = \frac{1}{2}$ (m=100, u=100) ir $\frac{m}{n} = \frac{3}{4}$ (m=150, u=50). Skaičiuojame $q = 0{,}1,\ 0{,}25,\ 0{,}50,\ 0{,}75,\ 0{,}90$ lygio kvantilio įverčius naudodami tiesioginį tradicinį kvantilio įvertinį $\widehat{K}_{yq} = \widehat{F}_y^{-1}(q)$, tiesioginį santykinį kvantilio įvertinį $\widehat{K}_{yq}^{sant} = \widehat{F}_y^{sant\ -1}(q)$ su koeficientu $\lambda = 0{,}5$, netiesioginį santykinį kvantilio įvertinį $\widehat{K}_{yq}^{sant*}(z)$ (1.26) su koeficientu $\omega = 0{,}5$. Naudojant kiekvieną iš įvertinių apskaičiuojami skirtingų q lygio kvantilio įverčiai. Taikant kiekvieną įvertinį gauname 1000 įverčių. Kvantilio įvertinių $\hat{\theta}_{yq} = \widehat{K}_{yq}, \widehat{K}_{yq}^{sant}, \widehat{K}_{yq}^{sant*}$ tikslumui palyginti skaičiuojamas gautų įverčių santykinis poslinkis ir santykinis efektyvumas atitinkamai:

$$\mathrm{SPosl}(\hat{\theta}_{yq}) = \frac{1}{\theta_{yq}} \left[\frac{1}{B} \sum_{i=1}^{B} (\hat{\theta}_{yq}^{(i)} - \theta_{yq}) \right];$$

$$\mathrm{SE}(\hat{\theta}_{yq}) = \frac{\mathrm{SVKP}(\widehat{K}_{yq})}{\mathrm{SVKP}(\hat{\theta}_{yq})}, \quad \mathrm{SVKP}(\hat{\theta}_{yq}) = \frac{1}{\theta_{yq}} \sqrt{\left[\frac{1}{B} \sum_{i=1}^{B} (\hat{\theta}_{yq}^{(i)} - \theta_{yq})^2 \right]}.$$

Čia $\hat{\theta}_{yq}^{(i)}$ - i-asis įvertis apskaičiuotas naudojant vieną iš įvertinių.

Lentelėse 3.5 ir 3.6 pateikti matematinio modeliavimo rezultatai.

3.5 lentelė. Reali populiacija: Įvertinių santykinis poslinkis (SPosl)

Table 3.5. Real population: relative bias of estimators

Įvertinys $\hat{\theta}$	$q = 0{,}10$	$q = 0{,}25$	$q = 0{,}50$	$q = 0{,}75$	$q = 0{,}90$
$\frac{m}{n} = \frac{1}{4}$					
$\widehat{K}_{yq}$	0,0039	0,0078	0,0079	0,0042	0,0076
$\widehat{K}_{yq}^{sant}$	−0,0123	−0,0022	0,0012	0,0019	0,0031
$\widehat{K}_{yq}^{sant*}$	0,0080	0,0039	0,0059	0,0049	0,0092
$\frac{m}{n} = \frac{1}{2}$					
$\widehat{K}_{yq}$	0,0014	0,0001	0,0002	−0,0022	−0,0005
$\widehat{K}_{yq}^{sant}$	−0,0039	−0,0016	−0,0025	−0,0022	−0,0011
$\widehat{K}_{yq}^{sant*}$	0,0032	−0,0020	0,0019	−0,0035	0,0045
$\frac{m}{n} = \frac{3}{4}$					
$\widehat{K}_{yq}$	−0,0023	−0,0048	−0,0018	−0,0032	−0,0043
$\widehat{K}_{yq}^{sant}$	0,0010	−0,0050	0,0050	−0,0010	−0,0020
$\widehat{K}_{yq}^{sant*}$	0,0090	−0,0001	0,0060	0,0023	0,0023

Pastebėta, kad labai mažiems q lygio kvantiliams tradicinis imties planu pagrįstas įvertinys yra efektyvus taikyti. Abiejų sudėtinių santykinių kvantilio įvertinių efektyvumas priklauso nuo pasirenkamų imties dydžių antrajame

3.6 lentelė. Reali populiacija: Įvertinių santykinis efektyvumas (SE)

Table 3.6. Real population: relative efficiency of estimators

Įvertinys $\hat{\theta}$	$q = 0{,}10$	$q = 0{,}25$	$q = 0{,}50$	$q = 0{,}75$	$q = 0{,}90$
$\frac{m}{n} = \frac{1}{4}$					
$\widehat{K}_{yq}$	1,000	1,000	1,000	1,000	1,000
$\widehat{K}_{yq}^{sant}$	0,983	1,157	1,232	1,185	1,195
$\widehat{K}_{yq}^{sant*}$	0,990	1,230	1,232	1,136	1,210
$\frac{m}{n} = \frac{1}{2}$					
$\widehat{K}_{yq}$	1,000	1,000	1,000	1,000	1,000
$\widehat{K}_{yq}^{sant}$	0,971	1,070	1,062	1,046	1,065
$\widehat{K}_{yq}^{sant*}$	0,953	1,091	1,106	1,029	1,083
$\frac{m}{n} = \frac{3}{4}$					
$\widehat{K}_{yq}$	1,000	1,000	1,000	1,000	1,000
$\widehat{K}_{yq}^{sant}$	0,626	0,635	0,644	0,647	0,623
$\widehat{K}_{yq}^{sant*}$	0,539	0,688	0,684	0,658	0,660

ėmime. Lyginant tiesioginį santykinį kvantilio įvertinį su netiesioginiu santykiniu įvertiniu esminių skirtumų nepastebėta. Kadangi skaičiuoti yra paprasčiau netiesioginį santykinį q lygio kvantilio įvertinį, pranašumą galėtume teikti jam.

3.3. Kvantilio pasikliautinojo intervalo vertinimas

Kvantilio įvertinių tikslumui nustatyti dažnai naudojamas pasikliautinasis intervalas. Kvantilio pasikliautinieji intervalai gali būti vertinami įvairiai. Konstruojant baigtinės populiacijos tyrimo kintamojo q lygio kvantilio K_{yq} pasikliautinojo intervalo įvertinį dažnai daroma prielaida, kad įvertinio pasiskirstymas yra normalusis. Šiame skyriuje pasiūlysime baigtinės populiacijos tyrimo kintamojo skirstinio kvantilio K_{yq} (3.35) pasikliautinojo intervalo įvertinius sudaromus naudojant imčių perrinkimo metodus. Sudarytus įvertinius lyginsime empiriškai, atlikdami matematinį modeliavimą.

Tarkime nagrinėjama baigtinė populiacija $\mathcal{U} = \{1,2,\dots,N\}$ su tyrimo kintamuoju y, kurio reikšmės $\{y_1, y_2, \dots, y_N\}$. Padarę prielaidą, kad tyrimo kintamojo y skirstinio kvantilio įvertinio $\widehat{K}_{yq}$ pasiskirstymas populiacijoje yra normalusis, įvertinio tikslumui nustatyti galime naudoti normaliuoju skirstiniu pagrįstą (NSP) pasikliautinojo intervalo įvertinį:

$$\left(\widehat{F}^{-1}\left(q - K_{\alpha/2}^{\mathcal{N}(0,1)}\sqrt{\widehat{\mathrm{D}}(\widehat{F}(\widehat{K}_q))}\right), \widehat{F}^{-1}\left(q + K_{\alpha/2}^{\mathcal{N}(0,1)}\sqrt{\widehat{\mathrm{D}}(\widehat{F}(\widehat{K}_q))}\right)\right), \tag{3.38}$$

čia $K_{\alpha/2}^{\mathcal{N}(0,1)}$ yra $1-\alpha/2$ standartinio normalaus skirstinio kvantilis. Jei iš baigtinės populiacijos $\mathcal{U}$ renkama paprastoji atsitiktinė imtis $\mathfrak{s}$, tuomet pasiskirstymo funkcijos įvertinio dispersijos įvertinio išraiška yra žinoma (Sitter, Wu 2001)

$$\widehat{\mathrm{D}}(\widehat{F}(\widehat{K}_q)) = \left(1 - \frac{n}{N}\right)\frac{\hat{s}^2}{n}, \quad \hat{s}^2 = \frac{1}{n-1}\sum_{k\in\mathfrak{s}}\left(z_k(\widehat{K}_q) - \overline{z}\right)^2,$$

$$\overline{z} = \frac{1}{n}\sum_{l\in\mathfrak{s}} z_l(\widehat{K}_q), \quad z_l(\widehat{K}_q) = \begin{cases} 1, & \text{jeigu } y_l < \widehat{K}_q, \\ 0, & \text{priešingu atveju.} \end{cases}$$

Patikrinti prielaidą apie tai, kad įvertinio $\widehat{K}_{yq}$ skirstinys normalusis nėra lengva. Todėl dažnai, vertindami kvantilio pasikliautinuosius intervalus, elgiamės taip, lyg prielaida apie tai, kad kvantlio įvertinio $\widehat{K}_{yq}$ skirstinys normalusis, būtų teisinga. Toliau siūlomos procedūros paremtos imčių perrinkimo metodais populiacijos kvantilio K_{yq} pasikliautinajam intervalui vertinti.

I. *Tradicinės savirankos metodas (*TS*).*

Siūloma procedūra:

- Iš baigtinės populiacijos $\mathcal{U}$ renkama n dydžio papratoji atsitiktinė imtis.
- Iš išrinktos imties renkamas n dydžio paprastasis atsitiktinis poimtis (savirankos imtis). Pažymėkime to poimčio q lygio kvantilį $\widehat{K}_{yq}^{(1)}$. Procesą kartojame B kartų, taip nustatydami kvantilius $\widehat{K}_{yq}^{(1)}$, $\widehat{K}_{yq}^{(2)}$,..., $\widehat{K}_{yq}^{(B)}$.
- Iš šios kvantilių aibės paėmę $\alpha/2$ ir $1-\alpha/2$ lygio kvantilius ir pažymėję juo $\widehat{K}_{yq}^{\mathrm{TS}}(\alpha/2)$ ir $\widehat{K}_{yq}^{\mathrm{TS}}(1-\alpha/2)$ atitinkamai gauname baigtinės populiacijos q lygio kvantilio K_{yq}, $1-\alpha$ lygmens pasikliautinojo intervalo įvertį:

$$\left(\widehat{K}_{yq}^{\mathrm{TS}}(\alpha/2), \widehat{K}_{yq}^{\mathrm{TS}}(1-\alpha/2)\right).$$

II. *Mastelio pakeitimo savirankos metodas (*MPS*).*

Šis metodas labai panašus į prieš tai buvusį metodą, skirtumas tas, kad pasikliautinasis intervalas baigtinės populiacijos kvantiliui sudaromas transformuotoms tyrimo kintamojo reikšmėms. Siūloma procedūra:

- Iš baigtinės populiacijos $\mathcal{U}$ renkama n dydžio papratoji atsitiktinė imtis. Tyrimo kintamojo y reikšmes pažymėkime $\{y_1, y_2,...,y_n\}$.
- Iš išrinktos imties, renkamas m dydžio poimtis (savirankos imtis). Tyrimo kintamojo y reikšmes žymėsime $\{y_1^*, y_2^*,...,y_m^*\}$.
- Tyrimo kintamo y reikšmės transformuojamos pakeičiant mastelį

$$\widetilde{y}_i = \overline{y}+\sqrt{\left(1-\frac{n}{N}\right)\frac{m}{n-1}}(y_i^*-\overline{y}), \quad i = 1,2,...,m, \quad \overline{y} = \frac{1}{n}\sum_{l\in\mathfrak{s}} y_l.$$

- Pažymėkime poimčio q lygio kvantilį $\widehat{K}_{\widetilde{y}q}^{(1)}$. Procesą kartojame B kartų, taip nustatydami kvantilius $\widehat{K}_{\widetilde{y}q}^{(1)}$, $\widehat{K}_{\widetilde{y}q}^{(2)}$,..., $\widehat{K}_{\widetilde{y}q}^{(B)}$.
- Iš šios kvantilių aibės paėmę $\alpha/2$ ir $1-\alpha/2$ lygio kvantilius ir pažymėję

juos $\widehat{K}^{\mathrm{MPS}}_{\tilde{y}q}(\alpha/2)$ ir $\widehat{K}^{\mathrm{MPS}}_{\tilde{y}q}(1-\alpha/2)$ atitinkamai gauname baigtinės populiacijos q lygio kvantilio K_{yq}, $1-\alpha$ lygmens pasikliautinojo intervalo įvertį:

$$\left(\widehat{K}^{\mathrm{MPS}}_{\tilde{y}q}(\alpha/2), \widehat{K}^{\mathrm{MPS}}_{\tilde{y}q}(1-\alpha/2)\right).$$

III. *Visrakčio metodas* (V).

Siūloma procedūra:

- Kvantilių K_q įverčiai skaičiuojami iš imties išmetus i-tąjį stebėjimą, tai yra $\widehat{K}^{(i)}_{yq}$, $i = 1, \ldots, n$.
- Surandami n įverčiai: $\widehat{K}^{(1)}_{yq}, \widehat{K}^{(2)}_{yq}, \ldots, \widehat{K}^{(n)}_{yq}$.
- Iš šios kvantilių aibės paėmę $\alpha/2$ ir $1-\alpha/2$ lygio kvantilius ir pažymėję juos $\widehat{K}^{\mathrm{V}}_q(\alpha/2)$ ir $\widehat{K}^{\mathrm{V}}_q(1-\alpha/2)$ atitinkamai gauname baigtinės populiacijos q lygio kvantilio K_{yq}, $1-\alpha$ lygmens pasikliautinojo intervalo įvertį:

$$\left(\widehat{K}^{\mathrm{V}}_{yq}(\alpha/2), \widehat{K}^{\mathrm{V}}_{yq}(1-\alpha/2)\right).$$

3.3.1. Įvertinių tikslumo tyrimas atliekant matematinį modeliavimą

Eksperimentas atliktas su dirbtinai sukurta N = 1000 dydžio baigtine populiacija. Sumodeliuoti du skirtingi duomenų rinkiniai.

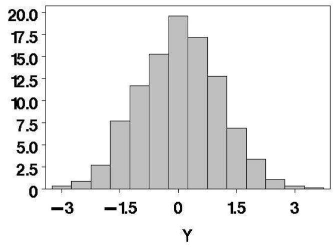
1 Atvejis. Generuojamos atsitiktinio dydžio reikšmės turinčios standartinį normalųjį pasiskirstymą su tankio funkcija

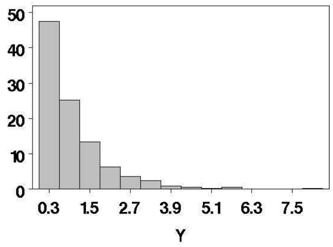
$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}, \quad -\infty < x < \infty.$$

2 Atvejis. Generuojamos atsitiktinio dydžio reikšmės turinčios eksponentinį pasiskirstymą su tankio funkcija

$$f(x) = \begin{cases} e^{-x}, & \text{jeigu } x \geq 0, \\ 0, & \text{priešingu atveju.} \end{cases}$$

Nagrinėjamų baigtinių populiacijų histogramos pateiktos 3.2 paveiksle.





3.2 pav. Baigtinių populiacijų histogramos, a) 1 Atvejis, b) 2 Atvejis
Fig. 3.2. Histograms of finite populations: a) Case 1, b) Case 2

Eksperimentui iš kiekvienos turimos baigtinės populiacijos buvo išrinkta 1000 paprastųjų atsitiktinių negrąžintinių $n = 200$ dydžio imčių. TS ir MPS atveju iš kiekvienos išrinktos imties papildomai buvo renkama $B = 1000$ paprastųjų atsitiktinių grąžintinių $m = 200$ imčių, dar kitaip vadinamų savirankos imčių. Parametrai, kuriems vertinamas pasikliautinasis intervalas yra tyrimo kintamojo skirtingų q lygių, atitinkamai 0,05, 0,25, 0,50, 0,75 ir 0,95 kvantiliai. Kai kurių tyrimo kintamojo kvantilio įvertinių pasiskirstymai pateikti 3.3 paveiksluose.

Eksperimentui pasikliautinųjų intervalų įverčiams skaičiuoti parenkamas pasikliovimo lygmuo $\alpha = 0{,}1$. Kiekvienai iš 1000 imčių skaičiuojami 90% pasikliautinieji intervalai ir nustatomas skaičius C tokių intervalų, kurie dengia tikrąją parametro reikšmę. Tikrasis pasikliovimo lygmuo vertinamas

$$\mathrm{PL} = \frac{C}{1000}.$$

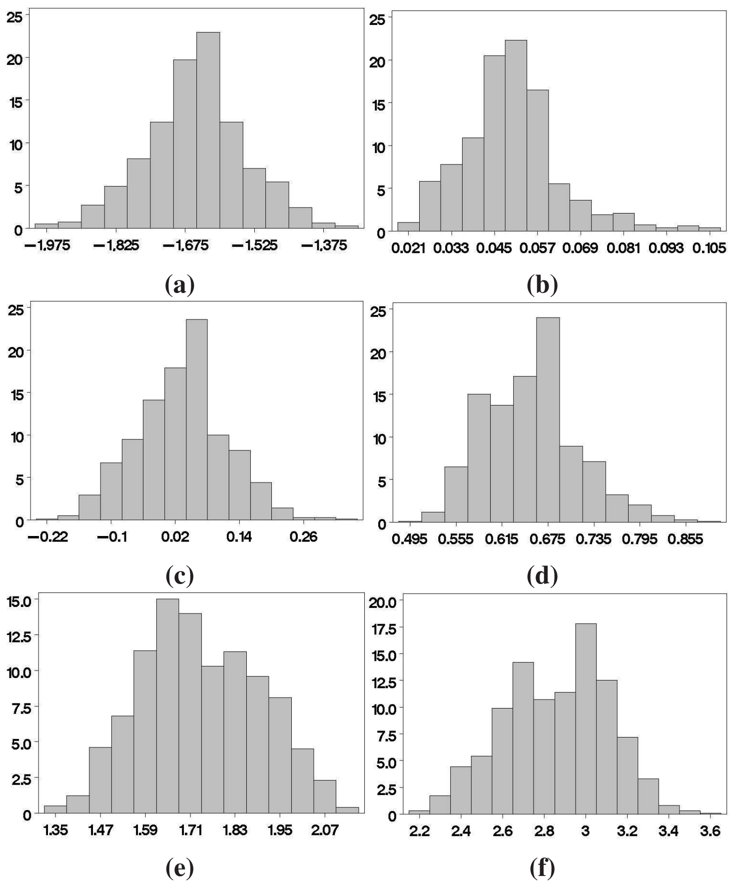
Be to apskaičiuojamas pasikliautinojo intervalo ilgių vidurkis

$$\bar{l} = \frac{1}{1000}\sum_{i=1}^{1000} l_i,$$

čia l_i pasikliautinojo intervalo ilgis nustatytas i-ajai imčiai.

Rezultatai pateikti 3.7 ir 3.8 lentelėse.

Modeliavimo rezultatai rodo, kad paprastosios atsitiktinės imties atveju, naudojant tiek normaliuoju skirstiniu pagrįstą, tiek tradicinį savirankos metodą, kvantilių pasikliautinųjų intervalų įverčiai gaunami panašūs. Tradicinės savirankos metodas, kaip alternatyva, gali būti naudojamas kvantilio pasikliau-



3.3 pav. 1000 kvantilio įverčių histograma kiekvienam iš atvejų, a) $\widehat{K}_{0,05}$ pasiskirstymas 1 Atveju, b) $\widehat{K}_{0,05}$ pasiskirstymas 2 Atveju, c) $\widehat{K}_{0,5}$ pasiskirstymas 1 Atveju, d) $\widehat{K}_{0,5}$ pasiskirstymas 2 Atveju, e) $\widehat{K}_{0,95}$ pasiskirstymas 1 Atveju, f) $\widehat{K}_{0,95}$ pasiskirstymas 2 Atveju

Fig. 3.3. Histograms of 1000 quantile estimates: a) Distribution of $\widehat{K}_{0.05}$ in Case 1, b) Distribution of $\widehat{K}_{0,05}$ in Case 2, c) Distribution of $\widehat{K}_{0,5}$ in Case 1, d) Distribution of $\widehat{K}_{0,5}$ in Case 2, e) Distribution of $\widehat{K}_{0,95}$ in Case 1, f) Distribution of $\widehat{K}_{0,95}$ in Case 2

3.7 lentelė. 90% pasikliautinųjų intervalų įverčių ilgių vidurkiai ($\bar{l}$)

Table 3.7. The average widths of 90% confidence intervals

Metodas	$K_{0,05}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,95}$
Dirbtinė populiacija (1 Atvejis)					
NSP	0,423	0,318	0,274	0,271	0,531
TS	0,409	0,345	0,302	0,298	0,529
MPS	0,367	0,345	0,302	0,298	0,529
V	0,020	0,017	0,013	0,015	0,028
Dirbtinė populiacija (2 Atvejis)					
NSP	0,046	0,108	0,194	0,347	0,901
TS	0,053	0,120	0,213	0,375	0,881
MPS	0,047	0,109	0,192	0,338	0,792
V	0,002	0,006	0,009	0,020	0,045

3.8 lentelė. 90% pasikliautinųjų intervalų pasikliovimo lygmens įverčiai (PL)

Table 3.7. Coverage of 90% confidence intervals

Metodas	$K_{0,05}$	$K_{0,25}$	$K_{0,50}$	$K_{0,75}$	$K_{0,95}$
Dirbtinė populiacija (1 Atvejis)					
NSP	0,918	0,882	0,885	0,891	0,887
TS	0,927	0,927	0,916	0,926	0,911
MPS	0,531	0,845	0,883	0,773	0,727
V	0,079	0,075	0,047	0,063	0,043
Dirbtinė populiacija (2 Atvejis)					
NSP	0,913	0,883	0,885	0,891	0,896
TS	0,928	0,917	0,913	0,925	0,904
MPS	0,000	0,000	0,876	0,856	0,852
V	0,071	0,075	0,047	0,063	0,090

tiniesiems intervalams konstruoti. Tokiu būdu sukonstruoti pasikliautinieji intervalai tenkina pasikliovimo lygmenį, nepriklausomai nuo kvantilio įvertinio pasiskirstymo, ir yra paprastai skaičiuojami.

Naudojant normaliuoju skirstiniu pagrįstą metodą kvantilių pasikliautinieji intervalai truputį per siauri, ir ne visada tenkina pasikliovimo lygmenį, tuo tarpu naudojant tradicinį savirankos metodą gauname ilgesnius intervalus, tačiau pastarieji tenkina pasikliovimo lygmenį. Naudojant visrakčio metodą gaunami labai siauri kvantilio pasikliautinieji intervalai. Pasiūlyta procedūra, naudojant visrakčio metodą, netinkama, reikia ieškoti kitų alternatyvų. Mastelio pakeitimo savirankos metodas geriausiai tinka vertinti medianos pasikliautinajam intervalui, tačiau jie taip pat gaunami per siauri.

3.4. Trečiojo skyriaus apibendrinimas

1. Sudaryti baigtinės populiacijos pasiskirstymo funkcijos sudėtiniai santykiniai įvertiniai. Išvestos tokių įvertinių apytikslių dispersijų išraiškos bet kokiam imties planui, tiek ir atskiru paprastosios atsitiktinės imties plano atveju.
2. Atliekant modeliavimą su realiais duomenimis autoriaus sudaryti baigtinės populiacijos pasiskirstymo funkcijos įvertiniai palyginti su tradiciniu nepaslinktuoju imties planu pagrįstu pasiskirstymo funkcijos įvertiniu.
3. Empiriškai palyginti baigtinės populiacijos tyrimo kintamojo skirstinio q lygio kvantilio tradicinis, tiesioginis santykinis ir netiesioginis santykinis įvertiniai naudojant dviejų ėmimų schemą.
4. Pasiūlytos imčių perrinkimo metodais paremtos procedūros kvantilio pasikliautinąjam intevalui vertinti. Tokie pasikliautinųjų intervalų įverčiai gali būti taikomi kai yra sudėtinga užrašyti kvantilio įvertinio dispersiją.

Bendrosios išvados

Išsprendus įvade suformuluotas problemas, gauti šie rezultatai:

1. Sudaryti sudėtiniai santykiniai baigtinės populiacijos sumos įvertiniai, naudojant iš ankstesnių tyrimų žinomą papildomą informaciją esant imties rotacijai. Modeliavimas atliktas su realiais gyventojų užimtumo tyrimo duomenimis. Atlikti eksperimentai parodė, kad tokie sudėtiniai santykiniai įvertiniai yra kur kas tikslesni lyginant su nepaslinktuoju baigtinės populiacijos sumos įvertiniu. Ieškant optimalių koeficientų būtina įtraukti kovariacijos narį, nuo to įvertinių tikslumas tik pagerėja. Realiam gyventojų užimtumo statistinio tyrimo imties planui tikslinga naudoti papildomą informaciją santykiniam įvertiniui visiems metų ketvirčiams, kuriems turime ankstesnių tyrimų duomenis.

2. Sudaryti sudėtinis regresinis ir santykinis baigtinės populiacijos pasiskirstymo funkcijos įvertiniai, o taip pat ir optimalūs įvertiniai su mažiausia dispersija, naudojant dviejų ėmimų schemą. Su realiais pajamų ir gyventojų sąlygų tyrimo duomenimis atliktas modeliavimas. Sudėtiniai pasiskirstymo funkcijos įvertiniai naudojantys papildomą informaciją iš ankstesnių rinkimų yra tikslesni už tradicinį imties planu pagrįstą baigtinės populiacijos pasiskirstymo funkcijos įvertinį. Tokių įvertinių efektyvumas priklauso nuo imties dydžių antrajame ėmime, o taip pat nuo kvantilio lygio.

3. Tiesioginis santykinis ir netiesioginis santykinis baigtinės populiacijos tyrimo kintamojo skirstinio kvantilio įvertinys empiriškai palygintas su tradiciniu imties planu pagrįstu kvantilio įvertiniu naudojant dviejų ėmimų schemą. Abu santykiniai įvertiniai gali būti tikslesni už tradicinį imties planu pagrįstą baigtinės populiacijos kvantilio įvertinį, tačiau tam įtakos turi ir parenkami imčių dydžiai antrajame ėmime. Kadangi netiesioginį santykinį kvantilio įvertinį skaičiuoti yra paprasčiau, pranašumą galėtume teikti jam.

4. Sukonstruoti baiginės populiacijos kvantilio pasikliautinojo intervalo įverčiai naudojant imčių perrinkimo procedūras. Tradicinės savirankos ir normaliuoju skirstiniu pagrįstu metodais konstruotų pasikliautinųjų intervalų įverčiai yra panašūs. Pasiūlytos visrakčio ir mastelio pakeitimo savirankos metodais paremtos procedūros nėra tinkamos kvantilio pasikliautinajam intervalui vertinti.

Literatūros sąrašas

Caron, N.; Ravalet, P. 2000. Estimation dans les enquêtes répétées: Application â Íenquête emploi en continu, in *Technical Report 0005, Méthodologie Statistique, INSEE, Paris*.

Cassel, C. M.; Särndal, C. E.; Wretman, J. 1976. Some results on generalized difference estimation and generalized regression estimation for finite populations, *Biometrika* 63(3): 615–620.

Chambers, R. L.; Dorfman, A. H.; Hall, P. 1992. Properties of estimators of the finite population distribution functions, *Biometrika* 79(3): 577–582.

Chambers, R. L.; Dunstan, R. 1986. Estimating distribution functions from survey data, *Biometrika* 73(3): 597–604.

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

Deville, J. C.; Särndal, C. E. 1992. Calibration estimators in survey sampling, *Journal of the American Statistical Association* 87: 376–382.

DiCiccio, T. J.; Efron, B. 1996. Bootstrap confidence intervals (with discussion), *Statistical Science* 11: 189–228.

Durbin, J. 1959. A note on the application of Quenouille’s method of bias reduction to the estimator of ratio’s, *Biometrika* 46: 477–480.

Eckler, A R. 1955. Rotation sampling, *The Annals of Mathematical Statistics* 54: 664–685.

Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics* 7: 1–26.

Francisco, C. A.; Fuller W. A. 1959. Quantile estimation with a complex survey design, *The Annals of Statistics* 19: 454–469.

Fuller, W. A. 2003. *Estimation for multiple phase design*, In: Analysis of Survey Data, eds. Chambers R. L. and Skinner. John Wiley and Sons, Chichester, 307–322.

Gross, S. T. 1980. Median estimation in sample surveys, *Journal of the American Statistical Association* 181–184.

Hajek, J. 1960. Limiting distributions in simple random sampling from a finite population, *Publications of the Matheatical Institute of the Hungarians Academy of Sciences*, 5(3): 361–374.

Harms, T.; Duchesne, P. 2006. On calibration estimation for quantiles, *Survey Methodology*, 32: 37–52.

Horvitz, D.; Thompson, D. 2008. A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association* 78(381): 663–685.

Isaki, C.; Fuller, W. 1982. Survey design under the regression superpopulation model, *Journal of the American Statistical Association* 77(377): 89–96.

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

Krapavickaitė, D. 2002. Kvantilio vertinimas baigtinėje populiacijoje, *Lietuvos matematikos rinkinys, spec. numeris* T42: 541–547.

Krapavickaitė, D.; Plikusas, A. 2005. *Imčių teorijos pagrindai*. Vilnius: Technika. 311 p.

Kuk, A. Y. C. 1988. Estimation of distribution functions and medians under sampling with unequal probabilities, *Biometrika* 75: 97–103.

Kuk, A. Y. C. 1988. A kernel method for estimating finite population distribution functions using auxiliary information, *Biometrika* 80(2): 385–392.

Kuk, A. Y. C.; Mak, T. K. 1989. Median estimation in the presence of auxiliary information, *Journal of The Royal Statistical society* B51: 261–269.

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

Lahiri, B, D. 1951. A method of sample selection providing unbiased ratio estimates, in *Bulletin of the International Statistical Institute*, 33(3): 133–140.

Lohr, L. S. 1999. *Sampling: Design and Analysis*. Pacific Grove: Duxbury Press. 494 p.

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

McCarthy, P. J. 1993. Standard error and confidence interval estimation for the median, *Journal of the American Statistical Association* 9(3): 673–689.

Montoya, R. Y.; Rueda, M.; Arcos, A. 2008. Confidence intervals for quantiles estimation using Jackknife techniques, *Computational Statistics*, 25–45.

Orusild, T. 1998. Confidence intervals for functions of quantiles under finite population sampling, in *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Jurmala, Latvia*, 51–55.

Orusild, T. 1999. Confidence intervals for functions of quantiles under finite population sampling, in *Proceedings of the Workshop on Survey Sampling Theory and Methodology, held in Palanga, Lithuania*, 51–55.

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

Quenouille, M. H. 1949. Problems in plane sampling, *The Annals of Mathematical Statistics* 20 :355–375.

Rao, J. N. K. 1994. Estimating totals and distribution function using auxiliary information at the estimation stage, *Journal of Official Statistics* 10(2):153–165.

Rao, J. N. K.; Kovar, J. G.; Mantel, H. J. 1990. On estimating distribution functions and quantiles from survey data using auxiliary information, *Biometrika* 77(2): 365–375.

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

Rueda, M.; Arcos, A.; Martinez, M. D. 2003. Difference estimators of quantiles in finite populations, *Sociedad de Estadistica e Investigacion Operativa Test* 12: 481–496.

Rueda, M.; Martinez, S.; Martinez, H.; Arcos, A. 2007. Estimation of the distribution function with calibration methods, *Journal of Statistical Planning and Inference* 137: 435–448.

Rueda, M.; Arcos, A.; Martinez, M. D.; Roman, Y.; 2004. Some improved estimators of quantile using auxiliary information in sample surveys, *Computational statistics and data analysis* 45: 825–848.

Rueda, M.; Martinez, S.; Martinez, H.; Arcos, A. 2007. Calibration methods for estimating quantiles. *Metrika* 66: 355-371.

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

Särndal, C. E.; Swensson, B.; Wretman, J. 1989. The weighted residual technique for estimating the variance of the general regression estimator of the finite popoulation total, *Biometrika* 76: 527–537.

Särndal, C. E.; Swensson, B.; Wretman, J. 1992. *Sampling: Design and Analysis*. New York: Springer-Verlag. 694 p.

Sen, A. R. 1971. Increased precision in Canadian waterfowl harvest survey through successive sampling, *The Journal of Wildlife Management* 35: 664–668.

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

Shao, J.; Tu, D. 1995. *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Shao, J. 2003. Impact of the bootstrap on sample survey, *Statistical Science* 18: 191–198.

Shao, J.; Wu, C. F. J. 1989. A general theory for jackknife variance estimation, *The Annals of Statistics* 17: 1176–1197.

Sheather, S. J.; Marron, S. J. 1990. Kernel quantile estimators, *Journal of the American Statistical Association* 85: 410–416.

Silverman, B. W. 1986. Density estimation for statistics and data analysis, in *Monographs on Statistics and Applied Probability, London: Chapman and Hall*, 1–22.

Sitter, R. R.; Wu, C. 2001. A note on Woodruff confidence intervals for quantiles, *Statistics and Probability Letters*, 52: 353–358.

Stein, R. 1989. An introduction to bootstrap methods, *Sociological Methods Research*, 18: 243–291.

Tukey, J. W. 1958. Bias and confidence in not-quite large sample, *The Annals of Mathematical Statistics* 29: 614.

Wang, S.; Dorfman, A. H. 1996. A new estimator for the finite population distribution function, *Biometrika* 83(3): 639–652.

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

Wu, C.; Sitter, R. R. 2001. Variance estimation for the finite population distribution function with complete auxiliary information, *The Canadian Journal of Statistics* 29: 289–308.

Autoriaus publikacijos disertacijos tema

Recenzuojamuose mokslo leidiniuose

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

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

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

Chadyšas, V.; Krapavickaitė, D. 2008. Dirbančių gyventojų skaičiaus vertinimas esant imties rotacijai, *Lietuvos matematikos rinkinys. LMD darbai* 48/49: 288–293. ISSN 0132-2818.

Kituose mokslo leidiniuose

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

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

Viktoras CHADYŠAS

BAIGTINĖS POPULIACIJOS PARAMETRŲ
STATISTINIAI ĮVERTINIAI ESANT IMTIES ROTACIJAI

Daktaro disertacija
Fiziniai mokslai, matematika (01P)

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

Doctoral Dissertation
Physical Sciences, Mathematics (01P)

2009 12 17. 9,5 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto leidykla „Technika“,
Saulėtekio al. 11, LT-10223 Vilnius, *http://leidykla.vgtu.lt*
Spausdino UAB „Baltijos kopija“,
Kareivių g. 13B, LT-09109 Vilnius, *http://www.kopija.lt*