Osakogumite kitsendustega hinnang

Kaja Sõstra[1]
Eesti Statistikaamet

Sissejuhatus

Valikuuringute üheks oluliseks ülesandeks on osakogumite hindamine. Kasvanud on nõudmine usaldusväärsete ja kooskõlaliste hinnangute järele, mis omakorda on andnud tõuke vastava teooria arengule. Senistes teoreetilistes käsitlustes on suurt tähelepanu pööratud hinnangute täpsuse tõstmisele, seda eriti väikese valimimahuga osakogumite korral. Sellele teemale on pühendatud raamatuid (Rao, 2003), kirjutatud teadusartikleid (Lehtonen jt, 2003, 2005) ja läbi viidud suuri uurimisprojekte (EURAREA konsortsium, 2005).

Käesoleva töö põhiteemaks on teine tähtis probleem osakogumite hindamise valdkonnas, see on hinnangute kooskõlalisuse ehk ühilduvuse probleem. Osakogumite ja üldkogumi parameetrid on sageli omavahel seotud. Näiteks peavad osakogumite kogusummad summeeruma üldkogumi kogusummaks. Hinnangute jaoks pole need seosed aga sageli täidetud, st hinnangud pole ühilduvad.

Statistika tarbijad ei aktsepteeri mitteühilduvaid hinnanguid. Statistiliste andmete ühilduvus on samuti oluline Euroopa statistikasüsteemi kvaliteedikomponent (Euroopa Komisjon, 2005, põhimõte 14). Mitteühilduvuse põhjusteks on hinnangute juhuslikkus, mitteaditiivsus, erinevate hinnangumeetodite kasutamine, hinnangute pärinemine erinevatest uuringutest jm. Kirjeldatud probleem tekib nii suurte kui ka väikeste osakogumite korral.

Tavaliselt kasutatakse lihtsaid ad hoc meetodeid ühilduvuse probleemi lahendamiseks. Sageli pole teada selliselt saadud hinnangute dispersiooni valem. Kui soovitakse saavutada ühilduvus

[1] Autor kaitses doktoritöö matemaatilise statistika erialal 21.detsembril 2007.

erinevatel tasemetel (väikesed osakogumid, suuremad osakogumid, üldkogum), siis muutub probleem matemaatiliselt keeruliseks.

1. Kasutatud hinnangud ja valikudisiainid

Käesolevas töös kasutatakse disainipõhist lähenemist, kus hinnangute omadused (keskväärtus, dispersioon, kovariatsioon) on määratud valikudisainiga. Vaadeldud on kahte valikudisaini: lihtne juhuslik valik (SI) ja hüpergeomeetriline valik (HG). Nendest SI-valik on võrdsete kaasamistõenäosustega tagasipanekuta valikudisain ja HG-valik on ebavõrdsete kaasamistõenäosustega tagasipanekuga valikudisain. Mõlemad valikudisainid on laialdaselt kasutuses valikuuringutes.

Avaldiste esitamisel on kasutatud valikuvektori meetodit (Traat, 2000), mis võimaldab samaaegselt käsitleda tagasipanekuta ja tagasipanekuga valikudisaine. Seega on paljud töö tulemused üldisemad kui seni kirjanduses käsitletud, nad kehtivad mõlema disainitüübi jaoks.

Antud töös eeldatakse, et osakogumite valimimahud pole liiga väikesed, st spetsiaalseid väikese osakogumi hinnanguid pole siin vaadeldud. GR-hinnang vajab esialgseid hinnanguid oma konstruktsioonis.

On vaadeldud kahte esialgset hinnangut osakogumite jaoks: lineaarset ja suhtehinnangut. Lineaarne hinnang on kirjanduses tuntud ka Horvitz-Thompsoni nime all (seda tagasipanekuta disainide korral). Mõlemad hinnangud on otsesed osakogumi hinnangud selles mõttes, et kasutavad ainult neid uuritava tunnuse väärtusi, mis on mõõdetud vaadeldavas osakogumis.

Suhtehinnangus kasutatakse lisaks osakogumi kohta teadaolevat abitunnuse kogusummat. Töös on ära toodud lineaarse ja suhtehinnangu dispersioonide ja kovariatsioonide üldised valemid, kusjuures suhtehinnangu korral on need Taylori reaksarendusele baseeruvad ligikaudsed avaldised.

Üldistest valemitest on tuletatud dispersiooni ja kovariatsiooni valemid SI- ja HG-valikute jaoks.

2. Osakogumi hinnangud

Käsitletakse osakogumeid ja osakogumi kogusumma hindamist, kasutades lineaarset ja suhtehinnangut. Antud töö seisukohalt on oluline teada nende hinnangute kovariatsioonimaatriksit. Seepärast tuletatakse üldised dispersiooni ja kovariatsioonivalemid osakogumi hinnangute jaoks ja ka valemid SI- ja HG-valikudisainide korral.

Osakogumite hinnangute kovariatsiooni pole siiani kirjanduses väga põhjalikult käsitletud. Suhtehinnangu ja tagasipanekuta disainide korral on kovariatsiooniavaldis toodud Särndal (1992), teatud erijuhtudel on kovariatsioonimaatriksi hinnangut vaadeldud raamatus Lehtonen (1995).

Antud töös saadud valemid võimaldavad teha mitmeid huvitavaid järeldusi osakogumite hinnangute sõltuvuse kohta. Selgus, et teatud valikudisainide ja suhtehinnangu korral on osakogumite hinnangud mittekorreleeritud. Samade valikudisainide jaoks võrdub osakogumi ja üldkogumi suhtehinnangute kovariatsioon osakogumi hinnangu dispersiooniga.

Esitatud on ka kaks näidet väikese üldkogumi ja kahe osakogumi korral. Esitatud tulemused on üldise iseloomuga, mis võimaldab neid rakendada nii tagasipanekuga kui ka tagasipanekuta valikudisainide korral.

3. Osakogumite kitsendustega hinnang

Käesolevas töös kasutatakse raamatus Knottnerus (2003) välja pakutud üldist kitsendustega hinnangut (General Restriction estimator), lühidalt GR-hinnang. Olgu $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_k)'$ uuritav parameetrite vektor ja $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_k\right)'$ k-dimensionaalne nihketa hinnangute vektor. Tähistame $\hat{\boldsymbol{\theta}}$ kovariatsioonimaatriksi $\mathbf{V}$. Parameetrid peavad rahuldama järgmisi lineaarseid kitsendusi:

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{c}, \tag{1}$$

kus $\mathbf{R}$ on $r \times k$ maatriks ja $\mathbf{c}$ on r-dimensionaalne konstantide vektor.

Teoreem 3.1. (Knottnerus, 2003, p. 328-329) *Kitsendustega hinnang $\hat{\boldsymbol{\theta}}^{GR}$, mis rahuldab tingimusi (1), ja hinnangu kovariatsioonimaatriks on järgmised:*

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{GR} &= \hat{\boldsymbol{\theta}} + \mathbf{K}(\mathbf{c} - \mathbf{R}\hat{\boldsymbol{\theta}}),\\ \mathbf{V}^{GR} &\equiv Cov(\hat{\boldsymbol{\theta}}^{GR}) = (\mathbf{I}_k - \mathbf{K}\mathbf{R})\mathbf{V},\end{aligned}$$

kus

$$\mathbf{K} = \mathbf{V}\mathbf{R}'(\mathbf{R}\mathbf{V}\mathbf{R}')^{-1}.$$

Töös rakendatakse teoreemis 3.1 esitatud Knottneruse üldisi ideid osakogumite hindamise ülesandele. Tuletatakse osakogumite jaoks GR-hinnangud ja nende dispersioonide ning kovariatsioonide avaldised. Tuletatud hinnangutele kandub üle GR-hinnangu optimaalsuse omadus. Seega on saadud osakogumite hinnangud dispersiooni mõttes parimad võimalikud kõigi osakogumite hinnangute hulgas, mis baseeruvad samadel alghinnangutel ja rahuldavad samu kitsendusi. Töös vaadeldakse summeeruvuskitsendust, st et D osakogumi kogusummad $Y_1, Y_2, \ldots, Y_D$ peavad summeeruma üldkogumi kogusummaks $\sum_{\mathfrak{D}} Y_d = Y$. Teoreemis 3.2 esitatakse kitsendustega hinnangu ja selle dispersiooni avaldised eeldusel, et üldkogumi kogusumma Y on teada. Kitsendused (1) avalduvad kujul:

$$\mathbf{R}\boldsymbol{\theta} = \begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ \cdots \\ Y_d \\ \cdots \\ Y_D \end{pmatrix} = Y. \qquad (2)$$

Teoreem 3.2. *Olgu osakogumite esialgsed hinnangud $\hat{Y}_i$, $i \in \mathfrak{D}$ ja hinnangute kovariatsioonimaaktriks $\mathbf{V} = (V_{ij})$. Kitsendustega (2) hinnang osakogumi U_d jaoks, $\hat{Y}_d^{GR}$, ja selle dispersioon, V_{dd}^{GR} on* *järgmised:*

$$\begin{aligned}\hat{Y}_d^{GR} &= \hat{Y}_d + \frac{Y - \sum_{\mathfrak{D}} \hat{Y}_i}{\sum\sum_{\mathfrak{D}} V_{ij}} \cdot \sum_{i \in \mathfrak{D}} V_{di},\\ V_{dd}^{GR} &= V_{dd} - \frac{\left(\sum_{i \in \mathfrak{D}} V_{di}\right)^2}{\sum\sum_{\mathfrak{D}} V_{ij}}.\end{aligned}$$

Antud metoodika annab võimaluse hinnangute arvutamiseks, kui parameetrite vahel peavad kehtima etteantud seosed. Töös on tõestatud teoreemid üldkogumi ja osakogumite GR-hinnangute kujust ja vastavatest dispersioonidest ja kovariatsioonidest. Tulemused on esitatud kolme tähtsa juhu jaoks: teadaolev üldkogumi kogusumma, samast või mõnest teisest uuringust hinnatud üldkogumi kogusumma, hinnatud ja tinglikult fikseeritud üldkogumi kogusumma. Esialgsete hinnangutena on eeldatud lineaarset ja suhtehinnangut. Tuletatud on ka valemid SI- ja HG-valikudisainide jaoks.

Tuletatud GR-hinnangutel on oluline omadus, nad on minimaalse dispersiooniga kõigi teiste samu kitsendusi rahuldavate ja samu esialgseid hinnanguid kasutavate hinnangute hulgas. Osutus, et GR-hinnangu dispersioon (v.a. tingliku GR-hinnangu juht) ei ole suurem esialgse hinnangu dispersioonist. Seega üldjuhul on GR-hinnangul mitu head omadust: rahuldab kitsendusi ja on täpsem kui esialgne hinnang. Töös arendatakse diskussiooni, kuidas GR-hinnangu analüütilise kuju uurimine võimaldab konstrueerida teisi lihtsamaid (pole vaja teada alghinnangu kovariatsioonimaatriksit), kuid siiski optimaalsele lähedasi hinnanguid.

4. Simuleerimisülesanne

Simuleerimiseksperimendi eesmärgiks oli uurida kitsendustega hinnangu käitumist praktikas ja seda, kuidas asümptootilised tulemused töötavad lõpliku valimimahu korral. Selleks moodustati 2000 isikust üldkogum ja võeti 10000 valimit suurusega ligikaudu 200 isikut, kasutades SI- ja HG-valikudisaine. Eksperimentides vaadeldakse kitsendustega hinnangut kolmel erijuhul: teadaolev üldkogumi kogusumma, hinnatud üldkogumi kogusumma ja tinglik kitsendustega hinnang. Simuleerimiseksperiment kinnitas töös saadud teoreetilisi tulemusi. Isegi asümptootilised valemid töötasid hoolimata küllaltki väikesest valimimahust hästi. Eksperiment illustreeris järgmisi aspekte:

- GR-hinnangud olid nihketa või oli nihe väga väike.
- GR-hinnangu dispersioon oli väiksem esialgse hinnangu dispersioonist. Suurim suhteline vähenemine oli osakogumis, mille esialgse hinnangu dispersioon oli suurim.
- Hinnatud kovariatsioonimaatriksi kasutamine GR-hinnangus suurendas vähesel määral hinnangu dispersiooni.
- Tinglik GR-hinnang oli vähem efektiivne kui sama osakogumi esialgne hinnang.
- Osakogumite GR-hinnangute korrelatsioon oli tugevam kui vastavatel esialgsetel hinnangutel. Tugev sõltuvus ilmnes isegi juhul, kui esialgsed hinnangud ei olnud korreleeritud. GR-hinnangute korrelatsioon oli negatiivne.
- Tingliku GR-hinnangu korral, kui üldkogumi kogusumma hinnang fikseeritakse, tuleb selle juhuslikkust kindlasti arvestada osakogumi hinnangu dispersiooni valemis, et mitte alahinnata dispersiooni.

Lõpetuseks võib öelda, et töös tuletatud summeeruvuskitsendust rahuldavad osakogumite hinnangutel on mitmeid häid omadusi, nad on ühilduvad ja üldiselt täpsemad kui esialgsed hinnangud. Nad on praktikas rakendatavad.

Kirjandus

[1] EURAREA Consortium with K. Sõstra among the members (2004), *Enhancing Small Area Estimation Techniques to meet European Needs (EURAREA project) Final Reference Volume: Vol.* **1–3** https://www.statistics.gov.uk/eurarea/default.asp .

[2] European Commission (2005), *European Statistics Code of Practice.* http://epp.eurostat.ec.europa.eu/ .

[3] Knottnerus, P. (2003), *Sample Survey Theory. Some Pythagorean Perspectives.* New York: Springer.

[4] Lehtonen, R., Pahkinen, E. (1995), *Practical Methods for Design and Analysis of Complex Surveys.* Chichester: Wiley.

[5] Sõstra, K. (2004), Comparison of Small Area Estimation Methods: Simulation Study in EURAREA Project. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **8**, 243–252.

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

[7] Traat, I., Meister, K., Sõstra, K. (2001), Statistical inference in sampling theory. *Theory of Stochastic Processes*, vol. **7(23)**, 301–316.

[8] Traat, I. (2000), Sampling design as a multivariate distribution. *New trends in Probability and Statistics 5, Multivariate Statistics.* Vilnius, Utrecht: VSP/TEV, 195–208.

