

TARTU ÜLIKOOL  
MATEMAATIKA-INFORMAATIKATEADUSKOND  
MATEMAATILISE STATISTIKA INSTITUUT

Helis Puksand

Hindamine osakogumites summeeruvuse kitsenduse korral

Bakalaureusetöö

Juhendaja:  
Natalja Lepik, PhD

TARTU

2013

# Sisukord

<b>Sissejuhatus</b> .....	<b>4</b>
<b>1 Tähistused ja definitsioonid</b> .....	<b>6</b>
<b>2 Kitsenduse kasutamine osakogumis</b> .....	<b>7</b>
2.1 Kitsenduse vektorkuju .....	7
2.2 Levinud disainipõhised hinnangud .....	9
2.2.1 Lineaarne hinnang.....	9
2.2.2 Üldistatud regressioonihinnang.....	10
2.2.3 Suhtehinnang.....	11
2.3 Kitsenduse rakendamine disainipõhiste hinnangutele .....	13
2.3.1 Lineaarsed hinnangud osakogumites .....	13
2.3.2 Suhtehinnang.....	13
<b>3 Kitsendusega hinnang GR</b> .....	<b>14</b>
3.1 Probleemid hinnanguga GR.....	16
<b>4 Lisainfo kasutamine MSE asendamisel</b> .....	<b>17</b>
4.1 Ühikmaatriks .....	17
4.2 Osakogumi mahud .....	19
4.3 Valimimahud osakogumites .....	21
4.4 Osakogumite arv .....	23
<b>5 Praktiline näide</b> .....	<b>25</b>
5.1 Andmestiku kirjeldus.....	25
5.2 Lihtne juhuslik tagasipanekuta valik .....	27
5.3 Multinomiaalne valikudisain .....	29

<b>Kokkuvõte</b> .....	<b>31</b>
<b>Summary</b> .....	<b>32</b>
<b>Kirjanduse loetelu</b> .....	<b>34</b>
<b>Lisad</b> .....	<b>35</b>
Lisa 1. Programmikood .....	35

## Sissejuhatus

Käesolev bakalaureusetöö käsitleb kitsendustega hinnangut osakogumite korral. Statistika tarbijad vajavad usaldusväärseid näitajaid, seetõttu on praktikas väga oluline, et hinnangud vastaksid teatud kitsendustele, mis kehtivad üldkogumi parameetrite puhul. Osakogumite hinnangute korral on oluliseks piiranguks, et saadud hinnangud summeeruksid suurema osakogumi või üldkogumi summaks või selle hinnanguks.

Töö esimeses pooles tutvustatakse kitsendusi ja nendele vastavat hinnangut. Vaadeldava kitsendusega hinnangu korral kerkib üles probleem, et selle keskmise ruutvea või nihke hindamiseks on vaja teada esialgse hinnangu keskmist ruutviga ja nihet, kuid praktikas neid enamasti teada ei ole. Seetõttu otsitaksegi töö teises pooles võimalusi, kuidas võiks keskmise ruutvea maatriksit asendada ilma, et hinnangu täpsus palju kannataks. Viimasena katsetatakse saadud tulemusi reaalsete andmete peal.

Töö põhineb mitmel allikal. Esmalt nimetame doktoritöö Lepik (2011), mille sisuks on uurida kitsendustega hinnanguid, mis baseeruvad üldistatud regressioon- ja sünteetilistel hinnangutel, osakogumite korral. Käesoleva töö aluseks on üks nendest hinnangutest (GR2). Doktoritöös ei peatuta pikemalt probleemil, mille kohaselt on selle kitsendustega hinnangu keskmise ruutvea (MSE) ja nihke hindamiseks vaja teada esialgse hinnangu MSE-d ja nihet, kuid mida enamasti teada ei ole. Seepärast keskendutigi antud bakalaureusetöös esialgse MSE asendamisele. Ideed selle keskmise ruutvea asendamiseks pärinevad allikast Sõstra ja Traat (2009), kus on pakutud välja kolm erinevat osakogumi hinnangut, mis baseeruvad erinevatel korrigeerivatel konstantidel. Lisaks on proovitud konstante konkreetsetel andmetel ja võrreldud saadud osakogumite hinnangute täpsust. Käesolevas töös on kasutatud neid konstante, et konstrueerida maatrikseid, mis asendavad esialgsete hinnangute tundmatut MSE maatriksit. Lisaks on põhiallikaks ka doktoritöö Sõstra (2007), kus on kirjeldatud kitsendustega hinnangut osakogumite korral, kuid mis keskendub ainult nihketa alghinnangutele.

Autori ülesandeks käesoleva töö kirjutamisel on uue valdkonnaga tutvumine. Olulisim on uurida erinevate aktsepteeritavate hinnangute saamise meetodeid ning ühe konkreetse hinnangu praktikas rakendamisel tekkivaid probleeme. Töö koostaja omapoolseks panuseks on materjali kokkuvõtlik esitus, töös vaadeldavate erijuhtude kehtivuse näitamine ja nende kohta näidete toomine, samuti tulemuste katsetamine reaalsete andmete peal. Bakalaureusetöö

on kirjutatud tekstiõtlusprogrammiga *MS Word*, arvutuslik näide viidi läbi statistikapaketiga *R*.

Autor tänab Tervise Arengu Instituudi tervisestatistika osakonda andmete kasutamise loa ja juhendajat Natalja Lepikut arvukate selgituste, paranduste ning töö struktuuri puudutavate ideede eest.

# 1 Tähistused ja definitsioonid

Antud punktis toome ära töös kasutatavad mõisted ja tähistused.

**Definitsioon 1.1.** Parameetri  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)^T$  hinnangu  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_D)^T$  nihkeks nimetatakse elementide  $\hat{\theta}_i, i = 1, 2, \dots, d$ , nihete vektorit

$$\mathbf{b}(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}. \quad (1.1)$$

**Definitsioon 1.2.** Juhuslike hinnangute vektori  $\hat{\boldsymbol{\theta}}: N \times 1$  keskmise ruutvea maatriksiks nimetatakse  $N \times N$  maatriksit

$$\mathbf{MSE}(\hat{\boldsymbol{\theta}}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]. \quad (1.2)$$

Maatriksi peadiagonaalil asetsevad elemendid on ühedimensiooniliste hinnangute  $\hat{\theta}_i$  traditsioonilised keskmised ruutvead,  $E(\hat{\theta}_i - \theta_i)^2$ . Väljaspool diagonaal on nn ristruutkeskmised vead,  $E(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)$ .

Paneme tähele, et analoogiliselt ühemõõtmelise juhuga kehtib MSE maatriksi korral järgmine omadus:

$$\mathbf{MSE}(\hat{\boldsymbol{\theta}}) = \mathbf{V}(\hat{\boldsymbol{\theta}}) + \mathbf{b}(\hat{\boldsymbol{\theta}})[\mathbf{b}(\hat{\boldsymbol{\theta}})]^T. \quad (1.3)$$

**Definitsioon 1.3.** Kui iga  $N$ -mõõtmelise veeruvektori  $\mathbf{x} \neq \mathbf{0}$  ja ruutmaatriksite  $\mathbf{A}, \mathbf{B}: N \times N$  korral  $\mathbf{x}'(\mathbf{A} - \mathbf{B})\mathbf{x} > 0$ , siis Löwneri järjestamisseaduse kohaselt  $\mathbf{A} \geq \mathbf{B}$ . Kui  $\mathbf{A} \geq \mathbf{B}$ , siis  $a_{ii} \geq b_{ii}, \forall i$  korral. (Rao ja Rao, 1998, lk 508)

## 2 Kitsenduse kasutamine osakogumis

Kõikse uuringu läbiviimine on väga kulukas ning keeruline tegevus, seepärast kasutatakse enamasti uuringute tegemiseks valikuuringuid, mille eesmärgiks on anda informatsiooni tundmatute suuruste kohta üldkogumis. Üldkogumit tähistame edaspidi sümboliga  $U$ . Ükski uuring ei ole läbi viidud ideaalsetes tingimustes – esineb probleeme nii valmi võtmise kui hindamise etapil. Viimati nimetatud etapil tekkivate vigade parandamisele keskendub ka käesolev töö.

Nõudlus järjest usaldusväärsema ja täpsema statistika järele kasvab. Seetõttu soovivad statistika kasutajad üha enam, et valikuuringute hinnangud vastaksid teatud kriteeriumidele ehk kitsendustele. Antud juhul keskendume osakogumi hinnangutele. Osakogumiks nimetatakse üldkogumi  $U$  alamhulka  $U_d, U_d \subset U$ , kusjuures  $U_d \cap U_g = \emptyset$ , kui  $d \neq g$ . Osakogumite arvu tähistame  $D$ -ga. Selle objektid on sama tüüpi nagu üldkogumi omad ja see määratakse osakogumi identifikaatori järgi. (Lepik ja Traat, 2012, lk 20)

Osakogumite korral on oluline kitsendus, et saadud hinnangud summeeruksid üldkogumi summaks või selle hinnanguks. Sellest tulenevalt üritatakse leida andmetega kooskõlalist osakogumi hinnangut, mis ei oleks halvem kui algne mittekooskõlaline hinnang. Üks võimalus selleks on kasutada Knottneruse (2003) pakutud üldist kitsendustega hinnangut. Osakogumite jaoks on seda hinnangut uuritud töös Sõstra (2007). Seda tutvustame järgmises peatükis.

### 2.1 Kitsenduse vektorkuju

**Definitsioon 2.1.** Olgu  $\theta = (\theta_1, \theta_2, \dots, \theta_D)^T$  parameetrite vektor, mis rahuldab järgmist lineaarset kitsendust:

$$\mathbf{R}\theta = c, \tag{2.1}$$

kus  $\mathbf{R}$  on  $1 \times D$  reavektor ja  $c$  on teadaolev konstant. Maatriksit  $\mathbf{R}$  nimetame edaspidi kitsenduste maatriksiks või lihtsalt kitsenduseks.

**Näide 2.1.** Osakogumi summad  $t_y^d = \sum_{U_d} y_i$ , kus  $d = 1, 2, \dots, D$ , peavad summeeruma üldkogumi summaks  $t_y = \sum_U y_i$ , ehk  $\sum_{d=1}^D t_y^d = t_y$ .

Sel juhul saame kitsendusevõrrandi (2.1) komponendid esitada maatrikstähistuste kujul järgmiselt,

$$\mathbf{R} = (1, 1, \dots, 1), \boldsymbol{\theta} = (t_y^1, t_y^2, \dots, t_y^D)^T \text{ ja } c = t_y. \quad (2.2)$$

■

Järgnevalt näitame, et üldistust piiramata võime kirjutada kitsenduse (2.1) kujul:

$$\mathbf{R}\boldsymbol{\theta} = 0. \quad (2.3)$$

Kui kitsenduse (2.1) korral  $c \neq 0$ , siis on alati võimalik valida  $\boldsymbol{\theta}_0$  nii, et

$$0 = \mathbf{R}\boldsymbol{\theta} - c = \mathbf{R}\boldsymbol{\theta} - \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0).$$

Võime defineerida uue parameetri  $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$  ning saame  $\mathbf{R}\tilde{\boldsymbol{\theta}} = 0$ .

**Näide 2.2.** Olgu meil kolm osakogumit, siis  $\boldsymbol{\theta} = (t_y^1, t_y^2, t_y^3)^T$ ,  $\mathbf{R} = (1, 1, 1)$  ja kitsendus (2.1) on kujul:

$$\mathbf{R}\boldsymbol{\theta} = (1, 1, 1)(t_y^1, t_y^2, t_y^3)^T = \sum_{d=1}^3 t_y^d = t_y.$$

Võrrandi vasaku ja parema poole võime kirjutada kujul:

$$(t_y^1 + t_y^2 + t_y^3) - t_y = 0$$

ehk

$$\left(t_y^1 - \frac{1}{3}t_y\right) + \left(t_y^2 - \frac{1}{3}t_y\right) + \left(t_y^3 - \frac{1}{3}t_y\right) = 0.$$

Tähistame esimese liidetava  $\tilde{t}_y^1$ , teise  $\tilde{t}_y^2$  ja kolmanda  $\tilde{t}_y^3$ . Saimegi võrrandi (2.3) ümber parametrizeeritud kujul:

$$0 = \tilde{t}_y^1 + \tilde{t}_y^2 + \tilde{t}_y^3 = \mathbf{R}\tilde{\boldsymbol{\theta}},$$

kus  $\tilde{\boldsymbol{\theta}} = (\tilde{t}_y^1, \tilde{t}_y^2, \tilde{t}_y^3)^T$ .

■

Üldiselt võib üldkogumi summa ise olla hinnatav parameeter, kus hindamiseks kasutatakse sama valimit, mis osakogumite puhul. Sel juhul

$$\mathbf{R} = (1, 1, \dots, 1, -1), \boldsymbol{\theta} = (t_y^1, t_y^2, \dots, t_y^D, t_y)^T, c = 0.$$

Antud töös seda olukorda ei vaadelda. Siin töös eeldame, et  $t_y$  on teada (näiteks registrist või mõnest eelnevast uuringust hinnatud ning antud uuringus fikseeritud).

## 2.2 Levinud disainipõhised hinnangud

Lähenemist, kus hinnangu keskvärtust ja sellest tulenevalt ka kovariatsiooni defineeritakse valikudisaini abil, nimetatakse disainipõhiseks lähenemiseks. Disainipõhiste hinnangute omadused on määratud valikudisainiga. Juhuslikku vektorit  $\mathbf{I} = (I_1, I_2, \dots, I_N)^T$  nimetatakse valikuvektoriks, kus  $I_i$  on valikuindikaator ja näitab objekti  $i$  valikute arvu ( $i \in U$ ). Seejuures on  $E(I_i)$  valikuindikaatori keskvärtus ehk oodatav valikute arv. Valikudisainiks nimetatakse valikuvektori  $\mathbf{I}$  jaotust:

$$\mathbf{I} \sim p(\mathbf{k}), p(\mathbf{k}) = P(\mathbf{I} = \mathbf{k}), \sum_{\mathbf{k}} p(\mathbf{k}) = 1.$$

Valikuindikaatorite  $I_i, I_j$  kovariatsiooni tähistame  $\Delta_{ij}$  ja  $I_i$  dispersiooni tähistame  $\Delta_{ii}$ . Uuritava tunnuse kogusumma üldkogumis  $U$  on  $t_y = \sum_U y_i$ . Osakogumi  $U_d$  jaoks defineerime binaarse tunnuse  $z$ , kus

$$z_i^d = \begin{cases} 1, & \text{kui } i \in U_d; \\ 0, & \text{kui } i \notin U_d. \end{cases}$$

Sellest tulenevalt saame osakogumisse kuuluvad objektid võrrandi  $y_i^d = z_i^d y_i$  abil. Uuritava tunnuse kogusumma osakogumis on seega  $t_y^d = \sum_U z_i^d y_i$ . (Lepik ja Traat, 2012, lk 7, 25-26)

### 2.2.1 Lineaarne hinnang

Lineaarne nihketa hinnang kogusummale on kujul:

$$\hat{t}_y = \sum_U \frac{I_i y_i}{E(I_i)}$$

ning osakogumi kogusummale kujul:

$$\hat{t}_y^d = \sum_{U_d} \frac{I_i y_i}{E(I_i)} = \sum_U \frac{I_i y_i^d}{E(I_i)}. \quad (2.4)$$

Tagasipanekuta valikudisainide korral nimetatakse saadud hinnanguid Horvitz-Thompsoni hinnanguks või ka  $\pi$ -hinnanguks. Tagasipanekuga disainide korral aga Hansen-Hurwitz hinnanguks. (Lepik ja Traat, 2012, lk 16, 26)

## 2.2.2 Üldistatud regressioonihinnang

Üldistatud regressioonihinnangut kasutatakse lisainformatsiooni olemasolul eesmärgiga tõsta hinnangu täpsust lisainfo abil. Lisainformatsiooniks nimetatakse üldkogumi kohta teadaolevat (valikuuringust sõltumatut) informatsiooni. Need andmed võivad tulla registritest nii tunnuste kui ka agregeeritud summade näol. Tunnuseid, mille väärtused on teada iga objekti  $i$  kohta üldkogumist, nimetatakse abitunnusteks. Üldistatud regressiooni korral leitakse mudel, mis seob uuritavat tunnust seletavate tunnustega. Viimaseid saab valida olemasolevate abitunnuste hulgast. Mudeli parameetrid hinnatakse ja seejärel kasutatakse mudelit vaatlemata väärtuste prognoosimiseks. Kogusumma hinnang moodustatakse valimiväärtuste ja prognoosiväärtuste funktsioonina. (Traat, 2012, lk 31)

Vaatame esialgu üldistatud regressioonimudelit üldkogumis.

Olgu  $y_i$  uuritav tunnus ja  $\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ji} \end{pmatrix} : J \times 1$  abitunnuste vektor objektil  $i, i \in U$ . Teisendame

üldkogumi kogusummat  $t_y$  liites ja lahutades sellele  $\hat{y}_i$ :

$$t_y = \sum_U (y_i + \hat{y}_i - \hat{y}_i) = \sum_U \hat{y}_i + \sum_U (y_i - \hat{y}_i),$$

kus  $\hat{y}_i$  on teada iga  $i \in U$  korral, kuid  $y_i$  on teada üksnes valimis. Hindame nihketult saadud valemi teist liiget ja saame regressioonihinnangu kogusummale  $t_y$ :

$$\hat{t}_{GREG} = \sum_U \hat{y}_i + \sum_S \frac{(y_i - \hat{y}_i)}{E I_i}.$$

Olgu  $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{pmatrix}$  regressiooni kordajate vektor. Regressioonimudel is eeldatakse populatsiooni väärtusi  $y_i$ . Kui need väärtused oleks teada, siis saab  $\boldsymbol{\beta}$  hindamiseks kasutada üldistatud vähimruutude hinnangut:

$$\hat{\boldsymbol{\beta}} = \left( \sum_U \frac{\mathbf{x}_i \mathbf{x}_i^T}{\sigma_i^2} \right)^{-1} \sum_U \frac{\mathbf{x}_i y_i}{\sigma_i^2}; J \times 1.$$

Tähistame saadud vektori  $\mathbf{B}$ -ga. Näeme, et üldkogumi põhjal arvutatav suurus  $\mathbf{B}$  sisaldab kahte summat: maatriksite kogusummat  $T = \sum_U \frac{\mathbf{x}_i \mathbf{x}_i^T}{\sigma^2}$  ja vektorite kogusummat  $t_{xy} = \sum_U \frac{\mathbf{x}_i y_i}{\sigma^2}$ . Hinnates neid summasid nihketa saame hinnangu  $\mathbf{B}$ -le:

$$\hat{\mathbf{B}} = \left( \sum_S \frac{\mathbf{x}_i \mathbf{x}_i^T}{\sigma^2 E I_i} \right)^{-1} \sum_S \frac{\mathbf{x}_i y_i}{\sigma^2 E I_i}.$$

Vektor  $\hat{\mathbf{B}}$  on arvutatav valimist, seega saame arvutada prognoositud  $y$ -i väärtused  $\hat{y}_i = \mathbf{x}_i \hat{\mathbf{B}}$ ,  $i \in U$ , ja valimi jääkide hinnangud  $\hat{e}_i = y_i - \hat{y}_i$ , kus  $y_i$  on ainult valimis teada. Eelnevat kasutades saame:

$$\hat{t}_{GREG} = \sum_U \mathbf{x}_i \hat{\mathbf{B}} + \sum_S \frac{\hat{e}_i}{E I_i}. \quad (2.5)$$

Valemi (2.5) esimest liidetavat nimetatakse sünteetiliseks hinnanguks  $\sum_U \mathbf{x}_i \hat{\mathbf{B}} = \hat{t}_{SYN}$  ja kasutatakse sageli iseseisva hinnanguna, sest tal on väiksem varieeruvus kui regressioonhinnangul.

Kuna  $\sum_S \frac{\hat{e}_i}{E I_i}$  on GREG hinnangu nihke nii-öelda kompensatsioon, siis see tähendab, et sünteetilise hinnangu korral võib tekkida suur nihe. Mõningatel juhtudel võivad sünteetiline ja üldine regressioonhinnang kokku langeda.

### 2.2.3 Suhtehinnang

Antud töös kasutatakse GREG hinnangu erijuhtu – suhtehinnangut (Traat, 2012, lk 39). Suhtehinnang kasutab ühte abitunnust. Mudel, mis seob uuritavat tunnust  $y$  abitunnusega  $x$  kasutab järgmisi eelduseid:

- regressioonimudel on ilma vabaliikmeta;
- ühe abitunnusega;
- uuritava tunnuse dispersioon on proportsionaalne abitunnuste väärtustega.

**Definitsioon 2.2.** Suhtehinnanguks nimetatakse hinnangut

$$\hat{t}_{ratio} = \frac{t_x}{\hat{t}_x} \hat{t}_y,$$

kus  $\hat{t}_y = \sum_U \frac{I_i y_i}{E(I_i)}$ ,  $\hat{t}_x = \sum_U \frac{I_i x_i}{E(I_i)}$  on vastavalt lineaarsed hinnangud summadele  $t_y = \sum_U y_i$  ja  $t_x = \sum_U x_i$ .

Osakogumis saab leida suhtehinnangut  $\hat{t}_{ratio}^d$  järgmiselt:

$$\hat{t}_{ratio}^d = \frac{t_x^d}{\hat{t}_x^d} \hat{t}_y^d, \quad (2.6)$$

kus  $\hat{t}_y^d = \sum_s \frac{I_i}{E(I_i)} y_i^d$  on lineaarne hinnang uuritavale tunnusele  $y_i^d = \begin{cases} y_i, & i \in U_d \\ 0, & \text{mujal} \end{cases}$  osakogumis,  $t_x^d = \sum_{U_d} x_i$  on abitunnuse kogusumma osakogumis ning  $\hat{t}_x^d$  temale vastav lineaarne hinnang.

Kui registris  $X$ -tunnust ei ole, siis võetakse tavaliselt  $\mathbf{X} = (1, 1, \dots, 1)^T: N \times 1$ . Sel juhul  $t_x^d = N_d$  ja  $\hat{t}_x^d = \hat{N}_d$ , kus  $N_d$  on osakogumi summa ning  $\hat{N}_d$  selle hinnang.

Sõstra (2007, lk 31-32) põhjal saame valemi kovariatsiooni jaoks kahe osakogumi suhtehinnangu vahel.

**Tulemus 2.3.** Kovariatsioon kahe osakogumi suhtehinnangute  $\hat{t}_{ratio}^d$  ja  $\hat{t}_{ratio}^g$  vahel on

$$Cov(\hat{t}_{ratio}^d, \hat{t}_{ratio}^g) = \begin{cases} \sum \sum_U \Delta_{ij} \tilde{u}_i^d \tilde{u}_j^d, & \text{kui } d = g; \\ 0, & \text{kui } d \neq g, \end{cases} \quad (2.7)$$

kus  $\tilde{u}_i^d = \frac{y_i^d - R_d h_i^d}{E(I_i)}$ , kus  $h_i^d = 1$  ja  $R_d = \frac{\sum_{U_d} y_i^d}{N_d} \quad \forall i = 1, \dots, N, \quad \forall d = 1, \dots, D$  korral.  $\Delta_{ij} = Cov(I_i, I_j)$ .

Tulemus on asümptootiline, see tähendab tuletatud funktsiooni  $Cov(\hat{t}_{ratio}^d, \hat{t}_{ratio}^g)$  Taylori rea lineaarsest liikmest.

## 2.3 Kitsenduse rakendamine disainipõhiste hinnangutele

Vaatleme, kas kitsendus (2.1) on rahuldatud enamlevinud olukordades.

### 2.3.1 Lineaarsed hinnangud osakogumites

Osakogumite lineaarsetest hinnangutest saame hinnangute vektori  $\hat{\theta} = (\hat{t}_y^1, \hat{t}_y^2, \dots, \hat{t}_y^D)^T$ , kus  $\hat{t}_y^d$  on defineeritud valemiga (2.4). Soovime, et  $R\hat{\theta} = t_y$ , kus  $t_y$  on üldkogumi summa:

$$R\hat{\theta} = (1, 1, \dots, 1) \begin{pmatrix} \hat{t}_y^1 \\ \hat{t}_y^2 \\ \vdots \\ \hat{t}_y^D \end{pmatrix} = \sum_{i=1}^D \hat{t}_y^d = \sum_{i=1}^D \sum_{U_d} \frac{I_i y_i}{E(I_i)} = \sum_U \frac{I_i y_i}{E(I_i)} = \hat{t}_y \neq t_y.$$

Seega üldjuhul ei ole fikseeritud üldkogumi summa korral tingimus (2.1) osakogumite lineaarsete hinnangute jaoks täidetud.

### 2.3.2 Suhtehinnang

Suhtehinnangud osakogumites moodustavad hinnangute vektori

$$\hat{\theta} = (\hat{t}_{ratio}^1, \hat{t}_{ratio}^2, \dots, \hat{t}_{ratio}^D)^T$$

ja kitsendus (2.1) saab kuju:

$$R\hat{\theta} = (1, 1, \dots, 1) \begin{pmatrix} \hat{t}_{ratio}^1 \\ \hat{t}_{ratio}^2 \\ \vdots \\ \hat{t}_{ratio}^D \end{pmatrix} = \sum_{d=1}^D \hat{t}_{ratio}^d = \sum_{d=1}^D \frac{t_x^d}{\hat{t}_x^d} \hat{t}_y^d \neq t_y.$$

Seega tingimus (2.2) ei ole täidetud. Paneme tähele, et isegi  $\sum_{d=1}^D \frac{t_x^d}{\hat{t}_x^d} \hat{t}_y^d \neq t_y$ , mis oli täidetud lineaarsete hinnangute korral.

Järgmises peatükis tutvustame uut hinnangut, mis baseerub disainipõhistel hinnangutel, kuid mille korral on kitsendus täidetud.

### 3 Kitsendusega hinnang GR

Nägime, et eespool vaadeldud hinnangud ei vastanud kitsendusele (2.1). Samuti ei ole kitsendus rahuldatud, kui erinevates osakogumites on rakendatud eri tüüpi hinnanguid. Sooviksime siiski leida hinnangut, mis kitsendusele vastaks.

Töodes Knottnerus (2003) ja Söstra (2007) väljatöötatud kitsendustega hinnangud eeldavad, et algne hinnangute vektor on nihketa. Tihti aga osutub otstarbekamaks kasutada nihkega hinnanguid. Näiteks kasutatakse neid väikeste osakogumite korral. Seda eelkõige seetõttu, et tihti saavutatakse nihkega hinnangut kasutades väiksem hinnangu varieeruvus. Sellised hinnangud on näiteks üldistatud regressioonihinnangud (GREG), mis on nihketa ainult asümptootiliselt, ja sünteetiline hinnang (SYN), mille varieeruvus on küllalt väike, kuid nihe võib tugevasti kasvada väikeste osakogumite korral. (Lepik, 2011, lk 75-76)

Alljärgnevalt defineerime hinnangu, mis rahuldab kitsendust (2.3) ja mille keskmine ruutviga on Löwneri järjestuse mõttes täpsem kui algse hinnangu  $\hat{\theta}$  keskmine ruutviga.

**Definitsioon 3.1.** Üldistatud kitsendusega hinnang (*General Restricted estimator*) on defineeritud järgmiselt:

$$\hat{\theta}_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\theta}, \quad (3.1)$$

kus  $\mathbf{R} = (1, 1, \dots, 1)$ ,  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_D)^T$ ,  $\mathbb{I}$  on ühikmaatriks,  $\mathbf{K} = \mathbf{MR}^T(\mathbf{RMR}^T)^{-1}$  ja  $\mathbf{M} = \mathbf{MSE}(\hat{\theta})$ .

**Märkus 3.1.** Paneme tähele, et osakogumite korral

$$\mathbf{RMR}^T = (1, 1, \dots, 1) \cdot \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1} & m_{m2} & \cdots & m_{mn} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = p,$$

kus  $p$  on skalaar, ja

$$\mathbf{K} = \mathbf{MR}^T(\mathbf{RMR}^T)^{-1} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1} & m_{m2} & \cdots & m_{mn} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \frac{1}{p} =$$

$$= \begin{pmatrix} \sum_{i=1}^n m_{1i} \\ \sum_{i=1}^n m_{2i} \\ \vdots \\ \sum_{i=1}^n m_{mi} \end{pmatrix} \cdot \frac{1}{p} = \begin{pmatrix} \frac{\sum_{i=1}^n m_{1i}}{p} \\ \frac{\sum_{i=1}^n m_{2i}}{p} \\ \vdots \\ \frac{\sum_{i=1}^n m_{mi}}{p} \end{pmatrix}.$$

Näitame, et hinnang  $\hat{\boldsymbol{\theta}}_{GR}$  rahuldab tõepoolest kitsendust (2.3). Arvestades, et

$$\mathbf{R}\mathbf{K} = \mathbf{R}\mathbf{M}\mathbf{R}^T (\mathbf{R}\mathbf{M}\mathbf{R}^T)^{-1} = \mathbf{I},$$

siis

$$\mathbf{R}\hat{\boldsymbol{\theta}}_{GR} = (\mathbf{R}\mathbf{I} - \mathbf{R}\mathbf{K}\mathbf{R})\hat{\boldsymbol{\theta}} = (\mathbf{R} - \mathbf{R})\hat{\boldsymbol{\theta}} = \mathbf{0}.$$

■

**Märkus 3.2.** Töös (Lepik, 2011, lk 39) on näidatud, et hinnangu  $\hat{\boldsymbol{\theta}}_{GR}$  nihe on

$$\mathbf{b}(\hat{\boldsymbol{\theta}}_{GR}) = (\mathbf{I} - \mathbf{K}\mathbf{R})\mathbf{b}, \quad (3.2)$$

keskmise ruutvea maatriks on

$$\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR}) = (\mathbf{I} - \mathbf{K}\mathbf{R})\mathbf{M} \quad (3.3)$$

ning Löwneri järjestuse mõttes kehtib ka

$$\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR}) \leq \mathbf{M}. \quad (3.4)$$

■

**Näide 3.1.** Olgu  $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$  vektor tundmatutest parameetritest,  $\mathbf{R} = (1, 1)$  ja  $c = 0$ .

Hinnangute vektori  $\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$  keskmise ruutveamaatriks olgu  $\mathbf{M}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 10 & 9 \\ 7 & 8 \end{pmatrix}$  ja nihe

$$\mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Valemi (3.3) kohaselt on  $\hat{\boldsymbol{\theta}}_{GR}$  keskmine ruutviga

$$\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{M} = \left( \mathbb{I} - \begin{pmatrix} 0,559 \\ 0,441 \end{pmatrix} (1, 1) \right) \begin{pmatrix} 10 & 9 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 0,5 & -0,5 \\ -0,5 & 0,5 \end{pmatrix},$$

mis tähendab, et  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}^{(1)}) = 0,5$  ja  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}^{(2)}) = 0,5$ . Näeme, et tõepoolest  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}^{(1)}) \leq \mathbf{M}(\widehat{\boldsymbol{\theta}}_1)$  ja  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}^{(2)}) \leq \mathbf{M}(\widehat{\boldsymbol{\theta}}_2)$ .

Valemi (3.2) kohaselt on  $\widehat{\boldsymbol{\theta}}_{GR}$  nihe

$$\mathbf{b}(\widehat{\boldsymbol{\theta}}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{b} = \left( \mathbb{I} - \begin{pmatrix} 0,559 \\ 0,441 \end{pmatrix} (1, 1) \right) \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -0,676 \\ 0,676 \end{pmatrix}.$$

■

### 3.1 Probleemid hinnanguga GR

Hinnangu GR korral on keskmise ruutvea ja nihke hindamiseks vaja teada esialgse hinnangu ruutkeskmist viga  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}})$  ja nihet  $\mathbf{b}$ , mida praktikas aga enamasti ei ole teada.

GR korral on keskmine ruutviga väiksem kui esialgse hinnangu korral,  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR}) \leq \mathbf{M}$ . Sama ei saa aga järeldada nihete korral, mis võib mõningatel juhtudel olla ka suurem kui esialgne nihe,  $\mathbf{b}(\widehat{\boldsymbol{\theta}}_{GR}) \not\leq \mathbf{b}$ .

**Märkus 3.3.** Valikuuringutes on enamik hinnanguid nihketa või asümptootiliselt nihketa. Seetõttu keskendume antud töös juhtudele, kus  $\mathbf{b} = 0$  ehk  $E\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$  ja  $\mathbf{M} = \mathbf{Cov}(\widehat{\boldsymbol{\theta}})$ .

## 4 Lisainfo kasutamine MSE asendamisel

Tavaliselt ei ole praktikas esialgse hinnangu MSE maatriks teada, kuid hinnangu GR korral oleks keskmise ruutvea hindamiseks seda siiski vaja. Seetõttu üritame järgnevalt asendada esialgse MSE maatriksi mõne teise maatriksiga, mis sisaldaks meile teadaolevat lisainformatsiooni.

Märkuse 3.3 kohaselt vaatleme asümptootiliselt nihketa esialgseid hinnanguid, mis tähendab, et

$$\mathbf{M} = \mathbf{Cov}(\hat{\boldsymbol{\theta}}) = \mathbf{V}. \quad (4.1)$$

Allpool uurime nelja erineva maatriksi, milles olemaid parameetreid me teame, sobivust esialgse hinnangu kovariatsiooni kohale. Ideed nende valiku kohta on pärit artiklist Sõstra ja Traat (2009).

### 4.1 Ühikmaatriks

Vaatame esmalt juhtu, kus kasutame GR hinnangu (3.1) defineerimisel  $\mathbf{V} = \mathbb{I}_D$ . Sel juhu saame:

$$\begin{aligned} \mathbf{K}_1 &= \mathbf{M}\mathbf{R}^T(\mathbf{R}\mathbf{M}\mathbf{R}^T)^{-1} = \mathbf{V}\mathbf{R}^T(\mathbf{R}\mathbf{V}\mathbf{R}^T)^{-1} = \mathbb{I}\mathbf{R}^T(\mathbf{R}\mathbb{I}\mathbf{R}^T)^{-1} = \\ &= \mathbf{R}^T(\mathbf{R}\mathbf{R}^T)^{-1} = \mathbf{R}^T \left( \sum_{i=1}^D \mathbf{1} \right)^{-1} = \frac{1}{D} \cdot \mathbf{R}^T \end{aligned}$$

ja hinnang (3.1) avaldub kujul:

$$\hat{\boldsymbol{\theta}}_{GR_1} = \left( \mathbb{I} - \frac{1}{D} \cdot \mathbf{R}^T \mathbf{R} \right) \hat{\boldsymbol{\theta}} = \left( \mathbb{I} - \begin{pmatrix} \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \\ \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \end{pmatrix} \right) \hat{\boldsymbol{\theta}} = \begin{pmatrix} 1 - \frac{1}{D} & -\frac{1}{D} & \dots & -\frac{1}{D} \\ -\frac{1}{D} & 1 - \frac{1}{D} & \dots & -\frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D} & -\frac{1}{D} & \dots & 1 - \frac{1}{D} \end{pmatrix} \hat{\boldsymbol{\theta}}. \quad (4.2)$$

Näitame, et ühikmaatriksi korral vastab saadud hinnang kitsendusele (2.3):

$$\mathbf{R}\widehat{\boldsymbol{\theta}}_{GR_1} = \mathbf{R} \begin{pmatrix} 1 - \frac{1}{D} & -\frac{1}{D} & \cdots & -\frac{1}{D} \\ -\frac{1}{D} & 1 - \frac{1}{D} & \cdots & -\frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D} & -\frac{1}{D} & \cdots & 1 - \frac{1}{D} \end{pmatrix} \widehat{\boldsymbol{\theta}} = \left(1 - D \cdot \frac{1}{D}, 1 - D \cdot \frac{1}{D}, \dots, 1 - D \cdot \frac{1}{D}\right) \widehat{\boldsymbol{\theta}} = \mathbf{0}.$$

Valem (3.3) ei sobi saadud hinnangu MSE maatriksi leidmiseks, sest GR hinnangu kuju on teistsugune. MSE leidmiseks kasutame definitsiooni (1.2). Paneme tähele, et kasutades seost  $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}$  saame parameetri  $\boldsymbol{\theta}$  jaoks kasutada avaldist  $\boldsymbol{\theta} = (\mathbb{I} - \mathbf{K}_1\mathbf{R})\boldsymbol{\theta}$ . Sel juhul saame:

$$\begin{aligned} \mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_1}) &= E \left( (\mathbb{I} - \mathbf{K}_1\mathbf{R})\widehat{\boldsymbol{\theta}} - (\mathbb{I} - \mathbf{K}_1\mathbf{R})\boldsymbol{\theta} \right) \left( (\mathbb{I} - \mathbf{K}_1\mathbf{R})\widehat{\boldsymbol{\theta}} - (\mathbb{I} - \mathbf{K}_1\mathbf{R})\boldsymbol{\theta} \right)^T \\ &= E \left( (\mathbb{I} - \mathbf{K}_1\mathbf{R})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\mathbb{I} - \mathbf{K}_1\mathbf{R})^T \right) \\ &= (\mathbb{I} - \mathbf{K}_1\mathbf{R})E \left( (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right) (\mathbb{I} - \mathbf{K}_1\mathbf{R})^T \\ &= (\mathbb{I} - \mathbf{K}_1\mathbf{R})\mathbf{MSE}(\widehat{\boldsymbol{\theta}})(\mathbb{I} - \mathbf{K}_1\mathbf{R})^T. \end{aligned} \quad (4.3)$$

Kui esialgsete hinnangute vektor  $\widehat{\boldsymbol{\theta}}$  on nihketa, siis valem (4.3) avaldub järgmise seose kaudu:

$$\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_1}) = (\mathbb{I} - \mathbf{K}_1\mathbf{R})\mathbf{Cov}(\widehat{\boldsymbol{\theta}})(\mathbb{I} - \mathbf{K}_1\mathbf{R})^T \quad (4.4)$$

Seega valem (4.4) sõltub ikkagi algse hinnangu kovariatsioonist. Paneme tähele, et kuna esialgsed hinnangud on nihketa,  $\mathbf{b}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ , siis ka  $\widehat{\boldsymbol{\theta}}_{GR_1}$  on nihketa:

$$\begin{aligned} \mathbf{b}(\widehat{\boldsymbol{\theta}}_{GR_1}) &= E\widehat{\boldsymbol{\theta}}_{GR_1} - \boldsymbol{\theta} = \\ &= E[(\mathbb{I} - \mathbf{K}_1\mathbf{R})\widehat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \\ &= (\mathbb{I} - \mathbf{K}_1\mathbf{R})E\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \\ &= (\mathbb{I} - \mathbf{K}_1\mathbf{R})\boldsymbol{\theta} - \boldsymbol{\theta} = \\ &= \boldsymbol{\theta} - \mathbf{K}_1\mathbf{R}\boldsymbol{\theta} - \boldsymbol{\theta} = \mathbf{0}, \end{aligned}$$

sest  $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}$ . Järelikult

$$\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_1}) = \mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_1}).$$

Järgmisena toome näitena, kuidas avalduks  $\widehat{\boldsymbol{\theta}}_{GR_1}$  kovariatsioonimaatriks, juhul kui ühikmaatriks olekski algse hinnangu kovariatsioonimaatriks. Sel eeldusel saame valemist (4.4) valemis (3.3) defineeritud MSE:

$$\begin{aligned}
(\mathbb{I} - \mathbf{K}_1 \mathbf{R}) \mathbf{V} (\mathbb{I} - \mathbf{K}_1 \mathbf{R})^T &= (\mathbf{V} - \mathbf{K}_1 \mathbf{R} \mathbf{V}) (\mathbb{I} - \mathbf{R}^T \mathbf{K}_1^T) = \\
&= \mathbf{V} - \mathbf{K}_1 \mathbf{R} \mathbf{V} - \mathbf{V} \mathbf{R}^T \mathbf{K}_1^T + \mathbf{K}_1 \mathbf{R} \mathbf{V} \mathbf{R}^T \mathbf{K}_1^T = \\
&= \mathbf{V} - \mathbf{K}_1 \mathbf{R} \mathbf{V} - \mathbf{V} \mathbf{R}^T \mathbf{K}_1^T + \mathbf{V} \mathbf{R}^T (\mathbf{R} \mathbf{V} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{V} \mathbf{R}^T \mathbf{K}_1^T = \\
&= (\mathbb{I} - \mathbf{K}_1 \mathbf{R}) \mathbf{V}.
\end{aligned}$$

**Näide 4.1.** Valemi 3.3 põhjal saame, et MSE maatriks

$$\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_1}) = \begin{pmatrix} 1 - \frac{1}{D} & -\frac{1}{D} & \cdots & -\frac{1}{D} \\ -\frac{1}{D} & 1 - \frac{1}{D} & \cdots & -\frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D} & -\frac{1}{D} & \cdots & 1 - \frac{1}{D} \end{pmatrix} \mathbf{V} = \begin{pmatrix} 1 - \frac{1}{D} & -\frac{1}{D} & \cdots & -\frac{1}{D} \\ -\frac{1}{D} & 1 - \frac{1}{D} & \cdots & -\frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D} & -\frac{1}{D} & \cdots & 1 - \frac{1}{D} \end{pmatrix}. \quad (4.5)$$

Osakogumite arv  $D \geq 1$  ning seetõttu  $(1 - \frac{1}{D}) \leq 1$  ehk  $\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_1})$  peadiagonaali elemendid on väiksemad või võrdsed kui algse maatriksi – ühikmaatriksi – vastavad elemendid.

■

**Märkus 4.1.** Juhul kui  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}) = \mathbb{I}$ , siis  $\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_1})$  avaldub valemiga (4.5) ning nihketa  $\hat{\boldsymbol{\theta}}$  korral kehtib  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}_{GR_1}) \leq \mathbf{Cov}(\hat{\boldsymbol{\theta}})$ . Kui aga  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}) \neq \mathbb{I}$ , siis  $\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_1})$  avaldub valemiga (4.4) ja Löwneri järjestuse kohta järeldusi teha ei saa.

## 4.2 Osakogumi mahud

Järgmisena vaatame juhtu

$$\mathbf{V} = \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_D \end{pmatrix}, \quad (4.6)$$

kus  $N_i$  on osakogumi  $U_i$  maht,  $i = 1, 2, \dots, D$ . Sel juhu definitsioonist 3.1 saame

$$\mathbf{K}_2 = \mathbf{V} \mathbf{R}^T (\mathbf{R} \mathbf{V} \mathbf{R}^T)^{-1} = \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_D \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{\sum_{i=1}^D N_i} = \begin{pmatrix} \frac{N_1}{\sum_{i=1}^D N_i} \\ \frac{N_2}{\sum_{i=1}^D N_i} \\ \vdots \\ \frac{N_D}{\sum_{i=1}^D N_i} \end{pmatrix}$$

ja

$$\widehat{\boldsymbol{\theta}}_{GR_2} = \left( \mathbb{I} - \begin{pmatrix} \frac{N_1}{\sum_{i=1}^D N_i} \\ \frac{N_2}{\sum_{i=1}^D N_i} \\ \vdots \\ \frac{N_D}{\sum_{i=1}^D N_i} \end{pmatrix} (1, 1, \dots, 1) \right) \widehat{\boldsymbol{\theta}} = \begin{pmatrix} 1 - \frac{N_1}{\sum_{i=1}^D N_i} & -\frac{N_1}{\sum_{i=1}^D N_i} & \cdots & -\frac{N_1}{\sum_{i=1}^D N_i} \\ -\frac{N_2}{\sum_{i=1}^D N_i} & 1 - \frac{N_2}{\sum_{i=1}^D N_i} & \cdots & -\frac{N_2}{\sum_{i=1}^D N_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N_D}{\sum_{i=1}^D N_i} & -\frac{N_D}{\sum_{i=1}^D N_i} & \cdots & 1 - \frac{N_D}{\sum_{i=1}^D N_i} \end{pmatrix} \widehat{\boldsymbol{\theta}}.$$

Näitame, et algmaatriksi (4.6) korral vastab  $\widehat{\boldsymbol{\theta}}_{GR_2}$  kitsendusele 2.3:

$$\begin{aligned} \mathbf{R} \widehat{\boldsymbol{\theta}}_{GR_2} &= \mathbf{R} \begin{pmatrix} 1 - \frac{N_1}{\sum_{i=1}^D N_i} & -\frac{N_1}{\sum_{i=1}^D N_i} & \cdots & -\frac{N_1}{\sum_{i=1}^D N_i} \\ -\frac{N_2}{\sum_{i=1}^D N_i} & 1 - \frac{N_2}{\sum_{i=1}^D N_i} & \cdots & -\frac{N_2}{\sum_{i=1}^D N_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N_D}{\sum_{i=1}^D N_i} & -\frac{N_D}{\sum_{i=1}^D N_i} & \cdots & 1 - \frac{N_D}{\sum_{i=1}^D N_i} \end{pmatrix} \widehat{\boldsymbol{\theta}} \\ &= \left( 1 - \frac{\sum_{i=1}^D N_i}{\sum_{i=1}^D N_i}, 1 - \frac{\sum_{i=1}^D N_i}{\sum_{i=1}^D N_i}, \dots, 1 - \frac{\sum_{i=1}^D N_i}{\sum_{i=1}^D N_i} \right) \widehat{\boldsymbol{\theta}} = \mathbf{0}. \end{aligned}$$

Analoogiliselt valemiga (4.4) saab näidata, et hinnangu  $\widehat{\boldsymbol{\theta}}_{GR_2}$  MSE on ühtlasi võrdne tema kovariatsiooniga ning avaldub kujul:

$$\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_2}) = (\mathbb{I} - \mathbf{K}_2 \mathbf{R}) \mathbf{Cov}(\widehat{\boldsymbol{\theta}}) (\mathbb{I} - \mathbf{K}_2 \mathbf{R})^T. \quad (4.7)$$

Järgmisena toome näitena, kuidas avalduks  $\widehat{\boldsymbol{\theta}}_{GR_2}$  kovariatsioonimaatriks eeldusel, et maatriks (4.6) on algse hinnangu kovariatsioonimaatriks. Sel juhul saame kasutada valemit (3.3).

**Näide 4.2.** Valemi (3.3) põhjal saame, et kovariatsioonimaatriks

$$\begin{aligned}
\mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_2}) &= \left( \mathbb{I} - \begin{pmatrix} \frac{N_1}{\sum_{i=1}^D N_i} & \frac{N_1}{\sum_{i=1}^D N_i} & \cdots & \frac{N_1}{\sum_{i=1}^D N_i} \\ \frac{N_2}{\sum_{i=1}^D N_i} & \frac{N_2}{\sum_{i=1}^D N_i} & \cdots & \frac{N_2}{\sum_{i=1}^D N_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N_D}{\sum_{i=1}^D N_i} & \frac{N_D}{\sum_{i=1}^D N_i} & \cdots & \frac{N_D}{\sum_{i=1}^D N_i} \end{pmatrix} \right) \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_D \end{pmatrix} = \\
&= \begin{pmatrix} N_1 \left(1 - \frac{N_1}{\sum_{i=1}^D N_i}\right) & -\frac{N_1 N_2}{\sum_{i=1}^D N_i} & \cdots & -\frac{N_1 N_D}{\sum_{i=1}^D N_i} \\ -\frac{N_1 N_2}{\sum_{i=1}^D N_i} & N_2 \left(1 - \frac{N_2}{\sum_{i=1}^D N_i}\right) & \cdots & -\frac{N_2 N_D}{\sum_{i=1}^D N_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N_1 N_D}{\sum_{i=1}^D N_i} & -\frac{N_2 N_D}{\sum_{i=1}^D N_i} & \cdots & N_D \left(1 - \frac{N_D}{\sum_{i=1}^D N_i}\right) \end{pmatrix}. \tag{4.8}
\end{aligned}$$

Osakogumite maht  $N_d \geq 0$  ning  $\left(1 - \frac{N_d}{\sum_{i=1}^D N_i}\right) \leq 1 \forall d$  korral. Seega  $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_2})$  peadiagonaali elemendid on väiksemad või võrdsed algse maatriksi (4.6) vastavate elementidega võrreldes, mis tähendab, et kehtib  $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_1}^{(i)}) \leq \mathbf{Cov}(\widehat{\boldsymbol{\theta}}^{(i)})$ ,  $i = 1, 2, \dots, D$ .

■

**Märkus 4.2.** Juhul kui  $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_D \end{pmatrix}$ , siis  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_2})$  avaldub valemiga

(4.8) ning nihketa  $\widehat{\boldsymbol{\theta}}$  korral kehtib  $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_2}) \leq \mathbf{Cov}(\widehat{\boldsymbol{\theta}})$ . Kui aga

$\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) \neq \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_D \end{pmatrix}$ , siis  $\mathbf{MSE}(\widehat{\boldsymbol{\theta}}_{GR_2})$  avaldub valemiga (4.7) ja Löwneri

järjestuse kohta järeldusi teha ei saa.

### 4.3 Valimimahud osakogumites

Järgmisena vaatame juhtu, kui

$$\mathbf{V} = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_D \end{pmatrix}, \tag{4.9}$$

kus  $n_i$  on valimi maht osakogumis  $U_i$ ,  $i = 1, 2, \dots, D$ . Sarnaselt eelmisele variandile jõuame hinnangu kujul:

$$\hat{\boldsymbol{\theta}}_{GR_3} = \left( \mathbb{I} - \begin{pmatrix} \frac{n_1}{\sum_{i=1}^D n_i} \\ \frac{n_2}{\sum_{i=1}^D n_i} \\ \vdots \\ \frac{n_D}{\sum_{i=1}^D n_i} \end{pmatrix} (1, 1, \dots, 1) \right) \hat{\boldsymbol{\theta}} = \begin{pmatrix} 1 - \frac{n_1}{\sum_{i=1}^D n_i} & -\frac{n_1}{\sum_{i=1}^D n_i} & \dots & -\frac{n_1}{\sum_{i=1}^D n_i} \\ -\frac{n_2}{\sum_{i=1}^D n_i} & 1 - \frac{n_2}{\sum_{i=1}^D n_i} & \dots & -\frac{n_2}{\sum_{i=1}^D n_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_D}{\sum_{i=1}^D n_i} & -\frac{n_D}{\sum_{i=1}^D n_i} & \dots & 1 - \frac{n_D}{\sum_{i=1}^D n_i} \end{pmatrix} \hat{\boldsymbol{\theta}},$$

mille korral kitsendus (2.3) on samuti rahuldatud:

$$\begin{aligned} \mathbf{R}\hat{\boldsymbol{\theta}}_{GR_3} &= \mathbf{R} \begin{pmatrix} 1 - \frac{n_1}{\sum_{i=1}^D n_i} & -\frac{n_1}{\sum_{i=1}^D n_i} & \dots & -\frac{n_1}{\sum_{i=1}^D n_i} \\ -\frac{n_2}{\sum_{i=1}^D n_i} & 1 - \frac{n_2}{\sum_{i=1}^D n_i} & \dots & -\frac{n_2}{\sum_{i=1}^D n_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_D}{\sum_{i=1}^D n_i} & -\frac{n_D}{\sum_{i=1}^D n_i} & \dots & 1 - \frac{n_D}{\sum_{i=1}^D n_i} \end{pmatrix} \hat{\boldsymbol{\theta}} \\ &= \left( 1 - \frac{\sum_{i=1}^D n_i}{\sum_{i=1}^D n_i}, 1 - \frac{\sum_{i=1}^D n_i}{\sum_{i=1}^D n_i}, \dots, 1 - \frac{\sum_{i=1}^D n_i}{\sum_{i=1}^D n_i} \right) \hat{\boldsymbol{\theta}} = \mathbf{0}. \end{aligned}$$

Jällegi saab analoogiliselt valemiga (4.4) näidata, et hinnangu kovariatsioon avaldub kujul:

$$\mathbf{Cov}(\hat{\boldsymbol{\theta}}_{GR_3}) = (\mathbb{I} - \mathbf{K}_3 \mathbf{R}) \mathbf{Cov}(\hat{\boldsymbol{\theta}}) (\mathbb{I} - \mathbf{K}_3 \mathbf{R})^T. \quad (4.10)$$

Järgmisena toome näitena, kuidas avalduks  $\hat{\boldsymbol{\theta}}_{GR_3}$  kovariatsioonimaatriks, kui maatriks (4.9) on ka algse hinnangu kovariatsioonimaatriks. Sel juhul saame kasutada valemit (3.3).

**Näide 4.3.** Analoogiliselt hinnanguga  $\hat{\boldsymbol{\theta}}_{GR_2}$  saame hinnangu  $\hat{\boldsymbol{\theta}}_{GR_3}$  kovariatsioonimaatriksi kasutades valemit 3.3:

$$\mathbf{Cov}(\hat{\boldsymbol{\theta}}_{GR_3}) = \begin{pmatrix} n_1 \left( 1 - \frac{n_1}{\sum_{i=1}^D n_i} \right) & -\frac{n_1 n_2}{\sum_{i=1}^D n_i} & \dots & -\frac{n_1 n_D}{\sum_{i=1}^D n_i} \\ -\frac{n_1 n_2}{\sum_{i=1}^D n_i} & n_2 \left( 1 - \frac{n_2}{\sum_{i=1}^D n_i} \right) & \dots & -\frac{n_2 n_D}{\sum_{i=1}^D n_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_1 n_D}{\sum_{i=1}^D n_i} & -\frac{n_2 n_D}{\sum_{i=1}^D n_i} & \dots & n_D \left( 1 - \frac{n_D}{\sum_{i=1}^D n_i} \right) \end{pmatrix}. \quad (4.11)$$

Valimi maht osakogumites on  $n_d \geq 0$  ning  $(1 - \frac{n_d}{\sum_{i=1}^D n_i}) \leq 1 \forall d$  korral. Seega  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}_{GR_3})$  peadiagonaali elemendid on väiksemad või võrdsed kui maatriksi (4.9) vastavad elemendid.

■

**Märkus 4.3.** Juhul kui  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_D \end{pmatrix}$ , siis  $\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_3})$  avaldub valemiga

(4.11) ning nihketa  $\hat{\boldsymbol{\theta}}$  korral kehtib  $\mathbf{Cov}(\hat{\boldsymbol{\theta}}_{GR_3}) \leq \mathbf{Cov}(\hat{\boldsymbol{\theta}})$ . Kui aga

$\mathbf{Cov}(\hat{\boldsymbol{\theta}}) \neq \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_D \end{pmatrix}$ , siis  $\mathbf{MSE}(\hat{\boldsymbol{\theta}}_{GR_3})$  avaldub valemiga (4.10) ja Löwneri

järjestuse kohta järeldusi teha ei saa.

#### 4.4 Osakogumite arv

Järgmisena vaatame juhtu, kui

$$\mathbf{V} = \begin{pmatrix} \frac{1}{D} & 0 & \cdots & 0 \\ 0 & \frac{1}{D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{D} \end{pmatrix} : D \times D, \quad (4.12)$$

kus  $D$  on osakogumite arv. Sel juhu saame definitsioonist 3.1:

$$\mathbf{K}_4 = \mathbf{V}\mathbf{R}^T(\mathbf{R}\mathbf{V}\mathbf{R}^T)^{-1} = \begin{pmatrix} \frac{1}{D} & 0 & \cdots & 0 \\ 0 & \frac{1}{D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{D} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot 1 = \begin{pmatrix} \frac{1}{D} \\ \frac{1}{D} \\ \vdots \\ \frac{1}{D} \end{pmatrix}$$

ja

$$\widehat{\boldsymbol{\theta}}_{GR_4} = \left( \mathbb{I} - \begin{pmatrix} \frac{1}{D} \\ \frac{1}{D} \\ \vdots \\ \frac{1}{D} \end{pmatrix} (1, 1, \dots, 1) \right) \widehat{\boldsymbol{\theta}} = \left( \mathbb{I} - \begin{pmatrix} \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \\ \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{D} & \frac{1}{D} & \dots & \frac{1}{D} \end{pmatrix} \right) \widehat{\boldsymbol{\theta}}.$$

Paneme tähele, et hinnang  $\widehat{\boldsymbol{\theta}}_{GR_4}$  võrdub hinnanguga  $\widehat{\boldsymbol{\theta}}_{GR_1}$  (valem (4.2)) ning seega ka kitsendus (2.3) on kujul  $\mathbf{R}\widehat{\boldsymbol{\theta}}_{GR_4} = 0$ .

Analoogiliselt valemiga (4.4) saab näidata, et hinnangu kovariatsioon avaldub kujul:

$$\mathbf{Cov}(\widehat{\boldsymbol{\theta}}_{GR_4}) = (\mathbb{I} - \mathbf{K}_4\mathbf{R})\mathbf{Cov}(\widehat{\boldsymbol{\theta}})(\mathbb{I} - \mathbf{K}_4\mathbf{R})^T.$$

## 5 Praktiline näide

### 5.1 Andmestiku kirjeldus

Käesolevas töös kasutame Eesti tervishoiutöötajate andmeid, mis on saadud Tervise Arengu Instituudist tervisestatistika osakonnast 2009. aasta märtsi seisuga. Andmed koguti statistilise aruande „Tervishoiutöötajate tunnipalk“ jaoks. Kokku on andmestikus 1112 tervishoiu asutuse, näiteks kliinikute, perearsti kabinettide, hambaarstikeskuste jne 21 785 meditsiinitöötajat, kelle kohta on mõõdetud 30 tunnust. Näite simuleerimiseks kasutame nendest järgmiseid:

- 1) vanus, mida kasutatakse valikutõenäosuste arvutamiseks;
- 2) maakond, mis jaotab meditsiinitöötajad osakogumiteks;
- 3) ametialakood, mis näitab meditsiinitöötaja eriala.

Üldkogumi jaotame maakonna põhjal 4 osakogumiks. Jaotuse aluseks on võetud Päästeameti piirkondade liigendus:

- 1) Põhja-Eesti, kuhu kuuluvad Harjumaa ja Tallinn;
- 2) Ida-Eesti, kuhu kuuluvad Ida-Virumaa ja Lääne-Virumaa;
- 3) Lääne-Eesti, kuhu kuuluvad Hiiumaa, Järvamaa, Läänemaa, Pärnumaa, Raplamaa ja Saaremaa;
- 4) Lõuna-Eesti, kuhu kuuluvad Jõgevamaa, Põlvamaa, Tartumaa, Tartu, Valgamaa, Viljandimaa ja Võrumaa.

Saadud osakogumite mahud on antud tabelis 5.1.

Uuritavaks tunnuseks on valitud meditsiiniõdede arv Eestis. Tunnuse ametialakood alusel moodustame binaarse tunnuse, kus 1 tähendab, et meditsiinitöötaja töötab õena, näiteks hambaravi õena, ämmaemandana või operatsiooniõena, ja 0 tähendab, et töötaja on mõne muu erialaga.

Valimi moodustamiseks kasutame kahte erinevat disaini: lihtsat juhuslikku tagasipanekuta valikut ja multinomiaalset valikudisaini. Lihtsa juhusliku tagasipanekuta valikudisaini korral

$$E(I_i) = \frac{n}{N} \text{ ja } \Delta_{ij} = \begin{cases} -\frac{n}{N} \left(1 - \frac{n}{N}\right) \frac{1}{N-1}, & \text{kui } i \neq j; \\ \frac{n}{N} \left(1 - \frac{n}{N}\right), & \text{kui } i = j. \end{cases}$$

Multinomiaalne disain on objekti suurusega võrdelise tõenäosusega disain, kus

$$E(I_i) = np_i, n \text{ on valimimaht, ja } \Delta_{ij} = \begin{cases} -np_i p_j, & \text{kui } i \neq j; \\ np_i(1 - p_i), & \text{kui } i = j. \end{cases}$$

Vajaliku tõenäosuse  $p_i, i = 1, 2, \dots, 21785$ , leiame töötajate vanuseid kasutades

$$p_i = \frac{\text{vanus}_i}{\sum_{i=1}^{21785} \text{vanus}_i}.$$

Praktilise näite eesmärgiks on uurida, kuidas töötavad peatükis 4 kirjeldatud hinnangud  $GR_1, \dots, GR_4$  ja võrrelda nende MSE-d esialgse hinnangu MSE-ga. Esialgse hinnangute vektori rolli võtame suhtehinnangu (definitsioon 2.2), mille arvutame kõigi valimite jaoks. Saadud hinnangute põhjal leiame hinnangud  $GR_1, \dots, GR_4$ .

GR hinnangute täpsuse võrdlemiseks omavahel ja ka esialgse vektoriga  $\hat{\theta}$  kasutame keskmist empiirilist ruutkeskmist viga, mis avaldub kujul:

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{t}_i^d - t^d)^2, \quad (5.1)$$

kus  $m$  on valimite arv,  $\hat{t}_i^d$  on  $d$ -nda osakogumi kogusumma valimihinnang  $i$ -nda valimi korral ning  $t^d$  on  $d$ -nda osakogumi tegelik kogusumma. Suhtehinnangute täpsuse leidmiseks kasutame valemit (2.7). Mida väiksemad MSE väärtused, seda paremini sobivad kasutatud hinnangud reaalse väärtustega.

Tabel 5.1. Osakogumite ja valimite mahud

Osakogumid	Osakogumi maht	Keskmise valimimaht osakogumis	
		lihtsa juhusliku valiku korral	multinomiaaldisaini korral
1	9484	217,8	215,3
2	2754	63,2	64,7
3	2682	61,4	65,6
4	6865	157,6	154,4
Kokku	21785	500	500

Hinnangute uurimiseks koostame 5000 valimit kogumahuga 500. Nimetatud disainide kasutamisel saadud osakogumite keskmised valimimahud on toodud tabelis 5.1. Valimite korral ei jäänud ükski osakogum tühjaks.

## 5.2 Lihtne juhuslik tagasipanekuta valik

Peatükis 4 näitasime, et kõik neli GR hinnangut rahuldavad kitsendust (2.3). Nüüd näitame, kuidas kasutatud asendusmaatriksid töötavad lihtsa juhusliku tagasipanekuta valikuga koostatud valimi peal ja seda just hinnangu täpsuse mõttes (valem 5.1). Lisaks leiame GR hinnangu (3.1), mis baseerub esialgsete hinnangute teoreetilisel MSE maatriksil (2.7).

Tabel 5.2. Õdede arvu hinnangute väärtused osakogumites, kitsenduse kehtimine ja tegelik arv üldkogumis lihtsa juhusliku valiku korral

Hinnang	Näide	Osakogum				$R \cdot (\theta - \hat{\theta})$
		1	2	3	4	
Suhtehinnang $\hat{\theta}$	1	2035,6	445,5	344,8	1574,1	600,9
	2	2266,9	526,5	498,1	1224,3	485,2
	3	2544,5	769,5	536,4	1836,5	-685,9
$\hat{\theta}_{GR}$	1	2284,4	533,2	417,7	1765,8	0,0
	2	2467,8	597,3	556,9	1379,0	0,0
	3	2260,6	669,5	453,2	1617,8	0,0
$\hat{\theta}_{GR_1}$	1	2185,8	595,7	495,1	1724,4	0,0
	2	2388,2	647,8	619,4	1345,6	0,0
	3	2373,0	598,0	364,9	1665,0	0,0
$\hat{\theta}_{GR_2}$	1	2297,2	521,5	418,8	1763,5	0,0
	2	2478,1	587,8	557,8	1377,2	0,0
	3	2245,9	682,8	452,0	1620,4	0,0
$\hat{\theta}_{GR_3}$	1	2297,6	521,2	418,1	1764,0	0,0
	2	2478,4	587,6	557,3	1377,6	0,0
	3	2245,4	683,1	452,7	1619,8	0,0
$\hat{\theta}_{GR_4}$	1	2185,8	595,7	495,1	1724,4	0,0
	2	2388,2	647,8	619,4	1345,6	0,0
	3	2373,0	598,0	364,9	1665,0	0,0
Tegelik		2128,0	699,0	578,0	1596,0	0,0

Suhtehinnang ja GR hinnang õdede arvule osakogumis, arvatuna viie erineva maatriksi põhjal, on toodud tabelis 5.2. Kõigi hinnangute jaoks on tabelis toodud kolm esimest simuleerimissammu (5000-st). Võrdluseks on lisatud ka tegelik õdede arv osakogumites. On näha, et hinnangud  $\hat{\theta}_{GR_2}$  ja  $\hat{\theta}_{GR_3}$  annavad väga sarnaseid hinnanguid ning  $\hat{\theta}_{GR_1}$  ja  $\hat{\theta}_{GR_4}$ , nagu mainitud peatükis 4, annavad tõesti täpselt sama tulemuse. Tabelist on ka näha, et erinevalt GR hinnangutest ei summeeru osakogumite suhtehinnangud üldkogumi summaks ja seega ei vasta kitsendusele (2.3).

Tabelis 5.3 on võrreldud hinnangute täpsust valemi (5.1) abil. Parema visualisatsiooni jaoks on MSE maatriksite peadiagonaali elemendid kujutatud vektoritena. Näeme, et kõikide MSE maatriksite peadiagonaali elemendid on väiksemad kui üldise MSE peadiagonaali elemendid ning võrratus (3.4) kehtib ka meie simulatsiooni korral ehk  $MSE(\hat{\theta}_{GR}) < MSE(\hat{\theta})$ .

Tabel 5.3. Hinnangute täpsus lihtsa juhusliku valiku korral

Hinnangute MSE	Osakogum			
	1	2	3	4
$MSE(\hat{\theta})$	81743,8	28803,0	23953,7	62967,1
$MSE(\hat{\theta}_{GR})$	55860,0	22633,5	17657,7	47721,7
$MSE(\hat{\theta}_{GR_1})$	64740,2	25659,9	22590,2	48632,8
$MSE(\hat{\theta}_{GR_2})$	55369,5	22473,2	17688,3	47730,1
$MSE(\hat{\theta}_{GR_3})$	55356,9	22471,2	17669,3	47727,7
$MSE(\hat{\theta}_{GR_4})$	64740,2	25659,9	22590,2	48632,8

Näeme, et asendusmaatriksite kasutamine hinnangutes  $GR_1, \dots, GR_4$  annab tulemusi, mis on lähedased hinnanguga GR. Lisaks nad ei sõltu esialgse hinnangu  $\hat{\theta}$  MSE maatriksist ning rahuldavad samuti kitsendust (2.3). Antud simulatsiooni korral on hinnangu  $GR_3$  täpsus koguni ligilähedaselt sama GR täpsusega. Seega tehtud simulatsiooni põhjal võiks lihtsa juhusliku tagasipanekuta valiku korral algse hinnangu MSE asendada peatükis 4 kirjeldatud maatriksitega, ilma et saadav hinnang või selle täpsus oluliselt kannataks.

### 5.3 Multinomiaalne valikudisain

Vaatleme, kuidas käituvad hinnangud, kui valimite moodustamiseks on kasutatud multinomiaalset valikudisaini.

Tabeli 5.4. Õdede arvu hinnangute väärtused osakogumites, kitsenduse kehtimine ja tegelik arv üldkogumis multinomiaalse valikudisaini korral

Hinnang	Näide	Osakogum				$R \cdot (\theta - \hat{\theta})$
		1	2	3	4	
Suhtehinnang $\hat{\theta}$	1	2248,2	719,0	638,5	1773,0	-377,7
	2	2109,5	747,4	669,1	1457,1	17,9
	3	2365,5	782,7	515,3	1793,3	-455,8
$\hat{\theta}_{GR}$	1	2098,7	657,1	590,1	1655,1	0,0
	2	2116,6	750,3	671,4	1462,7	0,0
	3	2185,1	708,0	456,9	1651,0	0,0
$\hat{\theta}_{GR_1}$	1	2153,8	624,6	544,1	1678,6	0,0
	2	2114,0	751,8	673,6	1461,6	0,0
	3	2251,6	668,8	401,3	1679,3	0,0
$\hat{\theta}_{GR_2}$	1	2083,8	671,2	592,0	1654,0	0,0
	2	2117,3	749,6	671,3	1462,8	0,0
	3	2167,1	725,1	459,2	1649,6	0,0
$\hat{\theta}_{GR_3}$	1	2085,8	669,9	588,7	1656,7	0,0
	2	2117,2	749,7	671,5	1462,6	0,0
	3	2169,5	723,5	455,1	1652,9	0,0
$\hat{\theta}_{GR_4}$	1	2153,8	624,6	544,1	1678,6	0,0
	2	2114,0	751,8	673,6	1461,6	0,0
	3	2251,6	668,8	401,3	1679,3	0,0
Tegelik		2128,0	699,0	578,0	1596,0	0,0

Tabelis 5.4 on toodud osakogumite kaupa suhtehinnang ja GR hinnangud õdede arvule, arvatuna viie erineva hinnangu põhjal.

Paneme tähele, et ka selle valikudisaini korral  $\hat{\theta}_{GR_2}$  ja  $\hat{\theta}_{GR_3}$  on üksteisele lähedased hinnangud ning  $\hat{\theta}_{GR_1}$  ja  $\hat{\theta}_{GR_4}$  annavad, nagu peatükis 4 märkisime, täpselt sama tulemuse. Näeme, nagu lihtsa juhusliku valiku korralgi, et erinevalt GR hinnangutest ei summeeru osakogumite suhtehinnangud üldkogumi summaks ja seega ei vasta kitsendusele (2.3). Tabelis on esitatud kolm esimest simuleerimissammu (5000-st). Võrdluseks on lisatud ka tegelik õdede arv osakogumites.

Tabelis 5.5 on kujutatud MSE maatriksite peadiagonaali elemendid. Võrreldud on hinnangute täpsust valemi (5.1) abil.

Tabel 5.5. Hinnangute täpsus multinomiaalse valikudisaini korral

Hinnangute MSE	Osakogum			
	1	2	3	4
$MSE(\hat{\theta})$	99709,2	41296,2	32259,0	78647,0
$MSE(\hat{\theta}_{GR})$	67626,1	23847,7	20020,0	62012,1
$MSE(\hat{\theta}_{GR_1})$	76600,7	28302,1	26133,0	64520,9
$MSE(\hat{\theta}_{GR_2})$	66741,2	22885,8	19904,7	61934,4
$MSE(\hat{\theta}_{GR_3})$	66822,2	22952,0	20119,3	62130,0
$MSE(\hat{\theta}_{GR_4})$	76600,7	28302,1	26133,0	64520,9

Tabelis 5.5 olevate keskmiste ruutvigade omavahelisel võrdlemisel on näha, et kõigi GR hinnangute MSE peadiagonaali elemendid on väiksemad kui üldise MSE maatriksi omad ning võrratus (3.4) kehtib ka meie simulatsiooni korral ehk  $MSE(\hat{\theta}_{GR}) < MSE(\hat{\theta})$ . Kõigi hinnangute MSE-d, peale  $MSE(\hat{\theta})$ , on sarnased. Seega antud simulatsiooni põhjal võiks ka multinomiaalse valikudisaini korral algse hinnangu MSE asendada peatükis 4 kirjeldatud maatriksitega, ilma et saadav hinnang või selle täpsus oluliselt kannataks.

## Kokkuvõte

Käesolevas töös tutvustasime kitsendusega üldist regressioonihinnangut osakogumite korral. Selleks tegime ülevaate kitsendusest, mille kohaselt saadud hinnangud peavad summeeruma üldkogumi summaks või selle hinnanguks. Tegime lühikese ülevaate nihkega hinnangutest ning uurisime üldist regressioonihinnangut.

Üldise regressioonihinnangu korral on probleemiks, et praktikas ei ole algse hinnangu keskmise ruutvea maatriks teada. Seepärast keskendusime töö teises pooles selle asendamisele mõne muu, teadaolevate parameetritega, maatriksiga. Kõigi proovitud maatriksite korral jäi nõutav kitsendus kehtima. Lisaks viisime läbi ka praktilise näite, et katsetada keskmise ruutvea maatriksi asendamisel saadud hinnanguid reaalsete andmete peal. Valimi moodustamiseks kasutati lihtsat juhuslikku tagasipanekuta valikut ning multinomiaalset valikudisaini.

Saadud hinnangud olid omavahel küllaltki lähedased. Erinevate GR hinnangute täpsuse vahel ei esinenud märkimisväärseid erinevusi ning nad kõik olid täpsemad kui suhtehinnang. Seega meie näite põhjal võib öelda, et proovitud maatriksid sobivad algse hinnangu MSE maatriksit asendama, ilma et hinnang või tema täpsus oluliselt halvemaks muutuks.

# **Estimation of Domains under Summation Restriction**

Bachelor thesis

Helis Puksand

## **Summary**

The purpose of this Bachelor thesis was to present domain estimator under summation restriction, and to try to replace initial estimator's mean square error matrix with different matrices.

Nowadays, there is an increasing demand for accurate statistics of domains. Useful indicators can be found from sample survey as well as different registers. The survey has to collect information together with domain identifiers. As a result, this brings about the problem that estimators do not sum up to the population totals. The last case is an important restriction for estimators.

In the first part of the thesis, restrictions and different estimators are presented, but these estimators do not satisfy the restriction the way we want it or do not satisfy it at all. That is why the general restriction estimator that corresponds to the restriction is presented in the second part of the work. With the observed estimator, we come across the problem that for estimating mean square error and bias, an initial estimator's mean square error and bias has to be known, but in real life it is usually unknown. Therefore, in the second part of the thesis we replace the initial estimator's mean square error matrix with different matrices which contain additional information that we already know. As a result, all these estimators that were composed with these matrices satisfied the summation restriction.

In addition, a practical experiment was carried out on real data to illustrate the theory about replacing GR estimator's initial mean square error matrix with different matrices. Samples were taken by two sampling designs, the simple random sampling without a replacement and the multinomial design. Both designs gave similar results.

The experiment confirmed that all GR estimators that we found using different matrices satisfy the summation restriction. Furthermore, their precisions were better than precision of the initial estimators and very similar to the real GR estimator that was composed with the

initial estimator's actual mean square error. Based on our experiment we could replace the initial estimator's mean square error matrix with tested matrices to calculate the GR estimator.

## Kirjanduse loetelu

- Knottnerus, P., 2003. Sample Survey Theory. Some Pythagorean Perspectives. Wiley, New York.
- Lepik, N., 2011. Estimation of domains under restrictions built upon generalized regression and synthetic estimators. Doctoral Dissertation. Tartu Ülikool.
- Lepik, N., Traat, I., 2012. Valikuuringute teooria I. Loengute konspekt. Tartu Ülikool.
- Päästeameti kodulehekülg. Kätte saadaval: <http://www.rescue.ee/> [Vaadatud 20.04.2013].
- Rao, C.R., Rao, M.B., 1998. Matrix algebra and Its Applications to Statistics and Econometrics. World Scientific, Singapur.
- Sõstra, K., 2007. Restriction estimation for domains. Doctoral Dissertation. Tartu.
- Sõstra, K., Traat, I., 2009. Optical domain estimation under summation restriction. Journal of Statistical Planning and Inference vol. 139, pp. 3928-3941
- Traat, I., 2012. Valikuuringute teooria edasijõudnutele. Loengute konspekt. Tartu Ülikool.

# Lisad

## Lisa 1. Programmikood

### Andmete korrastamine

```
#Andmete sisselugemine
andmed=read.table("C:/Kool/LÖPUTÖÖ/andmed.txt", sep = "\t", dec = ",", header=TRUE)
head(andmed)

#Üldkogumi maht
N=length(andmed[,6])

#Uuritav tunnus
y=andmed[,6]

#Õdede eraldamine:
ode=rep(0,N) #1, kui õde; 0 muujal
for (i in 1:N){
  if(y[i]==2230){ode[i]=1}
  if(y[i]==3221){ode[i]=1}
  if(y[i]==3232){ode[i]=1}
  if(y[i]>223001 & y[i]<223019){ode[i]=1}
  if(y[i]>323101 & y[i]<323119){ode[i]=1}
  if(y[i]==22300101){ode[i]=1}
  if(y[i]==32310101){ode[i]=1}
  if(y[i]==32320101){ode[i]=1}
}
y=ode

#Seletavad tunnused:
vanus=andmed$vanus
palk=andmed$kogutunnipalk
amet=andmed$ametialakood

#Osakogumiteks jagamine:
Maakond=andmed$Maakond
Osakogum=rep(1,N)
for (i in 1:N){
  if (Maakond[i]==37 | Maakond[i]==784){Osakogum[i]=1} #Põhja-Eesti
  if (Maakond[i]==44 | Maakond[i]==59){Osakogum[i]=2} #Ida-Eesti
  if (Maakond[i]==39 | Maakond[i]==51 | Maakond[i]==57 | Maakond[i]==67 |
    Maakond[i]==70 | Maakond[i]==74){Osakogum[i]=3} #Lääne-Eesti
  if (Maakond[i]==49 | Maakond[i]==65 | Maakond[i]==78 | Maakond[i]==795 |
    Maakond[i]==82 | Maakond[i]==84 | Maakond[i]==86){Osakogum[i]=4} #Lõuna-Eesti
}
table(Osakogum)

#Uus andmestik, kus on ainult simuleerimiseks vajalikud tunnused
vanus[is.na(vanus)==TRUE]=round(runif(1)*60)+1 #Tühjadele väärtustele omistame juhusliku vanuse
vanus[vanus==0]=round(runif(1)*60)+1
```

```

p=vanus/sum(vanus) #MN disaini jaoks
sum(p) #Kontroll, et summa =1
data=cbind(y,vanus,Osakogum, p)
head(data)

#Moodustame sellest uue Excel faili:
write.csv(data,"C:/Kool/LÕPUTÖÖ/uus_andmed.txt")

```

### Simulatsioon lihtsa juhusliku valiku korral

```

#Simulatsioon - LIHTNE JUHUSLIK VALIK
install.packages("sampling", lib="C:/Program Files/R/R-2.13.1/library")
library(sampling)

#-----ANDMED-----#
#Andmete sisse lugemine
andmed=read.table("C:/Kool/LÕPUTÖÖ/uus_andmed.txt", sep = ",", header=TRUE)
head(andmed)
N=length(andmed[,1])
D=4

#Osakogumite mahud
M=table(andmed$Osakogum)
N1=M[1]
N2=M[2]
N3=M[3]
N4=M[4]

#Vajalikud maatriksid
R=t(rep(1,4))
I=diag(4)

#-----MOODUSTAN VALIMID JA LEIAN VALIMIMAHU OSAKOGUMITES-----#
n=500
r=5000
f=n/N

n1=rep(NA,r)
n2=rep(NA,r)
n3=rep(NA,r)
n4=rep(NA,r)

#Osakogumisse kuulumine
d1=(Osakogum==1)*1 #Muudab 1 ja 0 jadaks
d2=(Osakogum==2)*1
d3=(Osakogum==3)*1
d4=(Osakogum==4)*1

#Osakogumis olevad õed üldkogumi pikkuses
yid1=d1*y
yid2=d2*y

```

```

yid3=d3*y
yid4=d4*y

summa1=rep(NA,r)
summa2=rep(NA,r)
summa3=rep(NA,r)
summa4=rep(NA,r)

#Valimid
for (i in 1:r){
  s=srswor(n,N) #SRS tagasipanekuta (1 - võetakse valimisse)
  s1=rep(0,N)
  s1=s*d1 #Valim esimeses osakogumis
  s2=rep(0,N)
  s2=s*d2
  s3=rep(0,N)
  s3=s*d3
  s4=rep(0,N)
  s4=s*d4

  #Osakogumi kogusumma hinnang
  summa1[i]=sum(s1*y/f)
  summa2[i]=sum(s2*y/f)
  summa3[i]=sum(s3*y/f)
  summa4[i]=sum(s4*y/f)

  #Hinnatav osakogumi maht
  N1_h=sum(s1/f)
  N2_h=sum(s2/f)
  N3_h=sum(s3/f)
  N4_h=sum(s4/f)

  #Valimimahud
  n1[i]=sum(s1)
  n2[i]=sum(s2)
  n3[i]=sum(s3)
  n4[i]=sum(s4)
}

#Maatriksis (tabelis) on keskmised, minimaalsed ja maksimaalsed valimimahud osakogumites
t=matrix(c("1",round(mean(n1),1),min(n1),max(n1),
          "2",round(mean(n2),1),min(n2),max(n2),
          "3",round(mean(n3),1),min(n3),max(n3),
          "4",round(mean(n4),1),min(n4),max(n4),
          "Population",round(n,1),min(n),max(n)),
dimnames=list(c("","","","",""),c("Domain","Mean","Minimum","Maximum")),
byrow=T,nrow=5)

#-----SUHTEHINNANGUD-----#
suht1_ode=summa1*N1/N1_h
suht2_ode=summa2*N2/N2_h
suht3_ode=summa3*N3/N3_h

```

```

suht4_ode=summa4*N4/N4_h
sum(suht1_ode[1],suht2_ode[1],suht3_ode[1],suht4_ode[1]) #Hinnatud summa

#Maatriksina kõik kokku
suhth_ode=t(rbind(suht1_ode[1:r], suht2_ode[1:r], suht3_ode[1:r], suht4_ode[1:r]))
sd(suhth_ode[,1])^2

suhth_ode[1:5,] #Vaatan esimest viit

#-----KOVARIATSIOONI 1. JUHT - ÜLDINE KOVARIATSIOON-----#
#Deltad
dij=f*(n-1)/(N-1)-f^2
dii=f*(1-f)

var1=function(yid,Nid){
  summa=0
  for (i in 1:N){
    abi=rep(dij, N)
    abi[i]=dii
    summa=summa+sum(abi*((yid[i]-sum(yid)/Nid)/f)*((yid-sum(yid)/Nid))/f)
  }
  return(summa)
}

cov=rep(NA,4)
cov[1]=var1(yid1, N1)
cov[2]=var1(yid2, N2)
cov[3]=var1(yid3, N3)
cov[4]=var1(yid4, N4)
V=diag(cov) #Teen maatriksiks, et hiljem kasutada

#-----GR HINNANGUTE LEIDMINE ISE VALITUD KOVARIATSIOONIDE KORRAL-----#
#ALGMAATRIKSID
#Esimene juht
V1=diag(4)
#Teine juht
diagonaal2=c(N1, N2, N3, N4)
V2=diag(diagonaal2)
#Kolmas juht
diagonaal3=c(round(mean(n1), digits = 0), round(mean(n2), digits = 0), round(mean(n3), digits = 0),
round(mean(n4), digits = 0)) #Keskised valimmahud
V3=diag(diagonaal3)
#Neljas juht
diagonaal4=rep(1/D,4)
V4=diag(diagonaal4)

#FUNKTSIOONID
#K arvutamine
K=function(V,R){
  RVRt=R%*%V%*%t(R)
  #a=svd(RVRt) #polegi vist vaja, sest mul tuleb alati RVRt arv
  #inverse=a$V%*%diag(1/a$d)%*%t(a$u) #ei tööta kui RVRt on lihtsalt arv

```

```

        #K=V%%t(R)%%inverse
        K=V%%t(R)*(RVRt[1])^(-1)
        return(K)
    }

print("K-d")
print("V-kontroll:")
K_k=K(V,R)
print("V1:")
K1=K(V1,R)
print("V2:")
K2=K(V2,R)
print("V3:")
K3=K(V3,R)
print("V4:")
K4=K(V4,R)

#GR arvutamine
GR=function(K,theta,R,l){
    GR=(I-K%%R)%%theta
    return(GR)
}

GR1=matrix(rep(NA,20000), byrow=T, nrow=r)
GR2=matrix(rep(NA,20000), byrow=T, nrow=r)
GR3=matrix(rep(NA,20000), byrow=T, nrow=r)
GR4=matrix(rep(NA,20000), byrow=T, nrow=r)
GR_k=matrix(rep(NA,20000), byrow=T, nrow=r)

for (i in 1:r){
    GR1[i,]=GR(K1,suhth_ode[i,]-1/4*sum(y),R,l)
    GR2[i,]=GR(K2,suhth_ode[i,]-1/4*sum(y),R,l)
    GR3[i,]=GR(K3,suhth_ode[i,]-1/4*sum(y),R,l)
    GR4[i,]=GR(K4,suhth_ode[i,]-1/4*sum(y),R,l)
    GR_k[i,]=GR(K_k,suhth_ode[i,]-1/4*sum(y),R,l)
}

print("GR-id")
print("V-kontroll")
GR_k[1:3,]+1/4*sum(y)
print("V1:")
GR1[1:3,]+1/4*sum(y)
print("V2:")
GR2[1:3,]+1/4*sum(y)
hist((GR2[,1])+1/4*sum(y))
print("V3:")
GR3[1:3,]+1/4*sum(y)
print("V4:")
GR4[1:3,]+1/4*sum(y)

#TEGELIK ÕDEDE ARV OSAKOGUMIS
sum(y*d1)

```

```

sum(y*d2)
sum(y*d3)
sum(y*d4)

#KITSENDUSE KONTROLL
kitsendus=function(GRid,R){
  kitsendus=R**GRid
  return(kitsendus)
}

#Suhtehinnangu korral
Rid=rep(NA,5) #Vaatan esimest viit
for (i in 1:5){
  Rid[i]=kitsendus(t(t(suhth_ode[i,])),R)
}

#GRid
Rid=rep(NA,5)#Vaatan esimest viit (praegu GR kontrollhinnangu korral)
for (i in 1:5){
  Rid[i]=kitsendus(t(t(GR_k[i,])),R)
}

#----- KOVARIATSIIONI 2. KUNI 6. JUHT - ISE VALITUD KOVARIATSIIONID-----#
V1=rep(NA,4)
V2=rep(NA,4)
V3=rep(NA,4)
V4=rep(NA,4)
V_k=rep(NA,4)

for (i in (1:4)){
  V_k[i]=var(GR_k[,i])
  V1[i]=var(GR1[,i])
  V2[i]=var(GR2[,i])
  V3[i]=var(GR3[,i])
  V4[i]=var(GR4[,i])
}

Simulatsioon multinomiaal disaini korral

#Simulatsioon - MULTINOMIAALJAOTUS
library(sampling)

#ANDMED
#Andmete sisselugemine
andmed=read.table("C:/Kool/LÖPUTÖÖ/uus_andmed.txt", sep = ",", header=TRUE)
head(andmed)
N=length(andmed[,1])
D=4

#Osakogumite mahud
M=table(andmed$Osakogum)
N1=M[1]

```

```

N2=M[2]
N3=M[3]
N4=M[4]

#Vajalikud maatriksid
R=t(rep(1,4))
I=diag(4)

#-----MOODUSTAN VALIMID JA LEIAN VALIMIMAHU OSAKOGUMITES-----#
n=500
r=5000
n1=rep(NA,r)
n2=rep(NA,r)
n3=rep(NA,r)
n4=rep(NA,r)

#Osakogumitesse kuulumine
d1=(Osakogum==1)*1 #Muudab 1 ja 0 jadaks
d2=(Osakogum==2)*1
d3=(Osakogum==3)*1
d4=(Osakogum==4)*1

#Osakogumisse kuuluvad õed üldkogumina
yid1=d1*y
yid2=d2*y
yid3=d3*y
yid4=d4*y

summa1=rep(NA,r)
summa2=rep(NA,r)
summa3=rep(NA,r)
summa4=rep(NA,r)

p1=p*n #Keskvärtus EI

#Valimid
for (i in 1:r){
  s=UPmultinomial(p1) #MN disain
  s1=rep(0,N)
  s1=s*d1 #Valim esimeses osakogumis (konspektis I)
  s2=rep(0,N)
  s2=s*d2
  s3=rep(0,N)
  s3=s*d3
  s4=rep(0,N)
  s4=s*d4

  #Osakogumi kogusumma hinnang
  summa1[i]=sum(s1*y/p1)
  summa2[i]=sum(s2*y/p1)
  summa3[i]=sum(s3*y/p1)
  summa4[i]=sum(s4*y/p1)
}

```

```

#Hinnatav osakogumi maht
N1_h=sum(s1/p1)
N2_h=sum(s2/p1)
N3_h=sum(s3/p1)
N4_h=sum(s4/p1)

#Valimimahud
n1[i]=sum(s1)
n2[i]=sum(s2)
n3[i]=sum(s3)
n4[i]=sum(s4)
}

#Maatriksis (tabelis) on keskmised, minimaalsed ja maksimaalsed valimimahud osakogumites
t=matrix(c("1",round(mean(n1),1),min(n1),max(n1),
          "2",round(mean(n2),1),min(n2),max(n2),
          "3",round(mean(n3),1),min(n3),max(n3),
          "4",round(mean(n4),1),min(n4),max(n4),
          "Population",round(n,1),min(n),max(n)),
dimnames=list(c("", "", "", "", ""),c("Domain","Mean","Minimum","Maximum")),
byrow=T,nrow=5)

#SUHTEHINNANGUD
suht1_ode=summa1*N1/N1_h
suht2_ode=summa2*N2/N2_h
suht3_ode=summa3*N3/N3_h
suht4_ode=summa4*N4/N4_h
sum(suht1_ode[1],suht2_ode[1],suht3_ode[1],suht4_ode[1]) #Hinnatud summa

#Maatriksina kõik kokku
suht_ode=t(rbind(suht1_ode[1:r], suht2_ode[1:r], suht3_ode[1:r], suht4_ode[1:r]))
suht_ode[1:5,] #Vaatan esimest viit

# KOVARIATSIOONI 1. JUHT - ÜLDINE KOVARIATSIOON
#Delta arvutamine
for (i in 1:N){
  dij=-n*p[i]*p
}

var1=function(yid,Nid){
  summa=0
  for (i in 1:N){
    abi=dij
    abi[i]=n*p[i]*(1-p[i])
    summa=summa+sum(abi*((yid[i]-sum(yid)/Nid)/(n*p[i]))*((yid-sum(yid)/Nid)/(n*p)))
  }
  return(summa)
}

cov=rep(NA,4)
cov[1]=var1(yid1, N1)

```

```

cov[2]=var1(yid2, N2)
cov[3]=var1(yid3, N3)
cov[4]=var1(yid4, N4)
V=diag(cov)

#-----GR HINNANGUTE LEIDMINE ISE VALITUD KOVARIATSIOONIDE KORRAL-----#
#ALGMAATRIKSID
#Esimene juht
V1=diag(4)
#Teine juht
diagonaal2=c(N1, N2, N3, N4)
V2=diag(diagonaal2)
#Kolmas juht
diagonaal3=c(round(mean(n1), digits = 0), round(mean(n2), digits = 0), round(mean(n3), digits = 0),
round(mean(n4), digits = 0)) #Keskmised valimimahud
V3=diag(diagonaal3)
#Neljas juht
diagonaal4=rep(1/D,4)
V4=diag(diagonaal4)

#FUNKTSIOONID
#K arvutamine
K=function(V,R){
  RVRt=R%*%V%*%t(R)
  #a=svd(RVRt) #polegi vist vaja, sest mul tuleb alati RVRt arv
  #inverse=a$V%*%diag(1/a$d)%*%t(a$u) #ei tööta kui RVRt on lihtsalt arv
  #K=V%*%t(R)%*%inverse
  K=V%*%t(R)*(RVRt[1])^(-1)
  return(K)
}

print("K-d")
print("V-kontroll:")
K_k=K(V,R)
print("V1:")
K1=K(V1,R)
print("V2:")
K2=K(V2,R)
print("V3:")
K3=K(V3,R)
print("V4:")
K4=K(V4,R)

#GR arvutamine
GR=function(K,theta,R,I){
  GR=(I-K%*%R)%*%theta
  return(GR)
}

GR1=matrix(rep(NA,20000), byrow=T, nrow=r)
GR2=matrix(rep(NA,20000), byrow=T, nrow=r)
GR3=matrix(rep(NA,20000), byrow=T, nrow=r)

```

```

GR4=matrix(rep(NA,20000), byrow=T, nrow=r)
GR_k=matrix(rep(NA,20000), byrow=T, nrow=r)

for (i in 1:r){
  GR1[i,]=GR(K1,suhth_ode[i,]-1/4*sum(y),R,l)
  GR2[i,]=GR(K2,suhth_ode[i,]-1/4*sum(y),R,l)
  GR3[i,]=GR(K3,suhth_ode[i,]-1/4*sum(y),R,l)
  GR4[i,]=GR(K4,suhth_ode[i,]-1/4*sum(y),R,l)
  GR_k[i,]=GR(K_k,suhth_ode[i,]-1/4*sum(y),R,l)
}

print("GR-id")
print("V-kontroll")
GR_k[1:3,]+1/4*sum(y)
print("V1:")
GR1[1:3,]+1/4*sum(y)
print("V2:")
GR2[1:3,]+1/4*sum(y)
print("V3:")
GR3[1:3,]+1/4*sum(y)
print("V4:")
GR4[1:3,]+1/4*sum(y)

#TEGELIK ÕDEDE ARV OSAKOGUMITES
sum(y*d1)
sum(y*d2)
sum(y*d3)
sum(y*d4)

#KITSENDUSE KONTROLL
kitsendus=function(GRid,R){
  kitsendus=R%%GRid
  return(kitsendus)
}

#Suhtehinnangu korral
Rid=rep(NA,5) #Vaatan esimest viit
for (i in 1:5){
  Rid[i]=kitsendus(t(t(suhth_ode[i,])),R)
}

#GRide korral
Rid=rep(NA,5) #Vaatan esimest viit (praegu kontroll GR korral)
for (i in 1:5){
  Rid[i]=kitsendus(t(t(GR_k[i,])),R)
}

#----- KOVARIATSIOONI 2. KUNI 6. JUHT - ISE VALITUD KOVARIATSIOONID-----#
V1=rep(NA,4)
V2=rep(NA,4)
V3=rep(NA,4)
V4=rep(NA,4)

```

```
V_k=rep(NA,4)

for (i in (1:4)){
  V_k[i]=var(GR_k[,i])
  V1[i]=var(GR1[,i])
  V2[i]=var(GR2[,i])
  V3[i]=var(GR3[,i])
  V4[i]=var(GR4[,i])
}
```

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina Helis Puksand

(sünnikuupäev: 01.11.1990)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Hindamine osakogumites summeeruvuse kitsenduse korral“, mille juhendaja on Natalja Lepik,
  - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **06.05.2013**