

**TARTU ÜLIKOOL**  
**Matemaatikateaduskond**  
**Matemaatilise Statistika Instituut**

**T. Kaart**

**DISPERSIOONIKOMPONENTIDE JA PÄRITAVUSKOEFIITSIENDI  
HINDAMINE LOOMAPOPOPULATSIOONIDES**

**Magistritöö**

**Juhendaja: dots. T. Möls**

**Tartu 1997**

## Sisukord

Sissejuhatus .....	3
1. Sissejuhatus aretusteooriasse .....	5
1.1. Lineaarne geneetiline mudel .....	5
1.2. Päritavuskoefitsient .....	7
2. Üldine meetodika dispersioonikomponentide ja päritavuskoefitsiendi hindamiseks ..	10
2.1. Üldine lineaarne mudel .....	10
2.2. Dispersioonikomponentide hindamine üldisest lineaarsest mudelist .....	12
3. Dispersioonikomponentide ja päritavuskoefitsiendi hindamine tasakaalulisel juhul ..	15
3.1. ANOVA hinnangud .....	15
3.2. Dispersioonikomponentide ja päritavuskoefitsiendi hinnangute varieeruvus .....	18
3.3. Päritavuskoefitsiendi illegaalse hinnangu saamise tõenäosus .....	23
4. Dispersioonikomponentide hindamine mittetasakaalulisel juhul .....	32
4.1. Hendersoni meetodid .....	32
4.2. MIVQUE meetod .....	35
4.3. MIVQUE(0) meetod .....	38
4.4. ML meetod .....	39
4.5. REML meetod .....	41
5. Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodi võrdlus dispersioonikomponentide ja päritavuskoefitsiendi hindamisel .....	44
Kokkuvõte .....	56
Kasutatud kirjandus .....	58
Estimation of variance components and heritability coefficient in animal populations. Abstract .....	60
Lisa. Programm dispersioonikomponentide ja päritavuskoefitsiendi hindamise modelleerimiseks mittetasakaalulisel juhul segamudeli korral .....	62

## Sissejuhatus

Läbi aegade on inimene püüelnud selle poole, et saada suurema jõudlusega ja vastupidavamaid koduloomi. Kiireim tee toodangu suurendamiseks on parandada söötmist ja pidamistingimusi. Kuid sellisel moel on võimalik saavutada vaid indiviidi potentsiaalse jõudluse maksimaalne avaldumine. „Halva genotüübiga” loom ei hakka ka ideaalsetes keskkonnatingimustes korralikult tootma. Samuti ei kajastu keskkonnatingimuste mõju oluliselt järglaste omadustes. Seega on pidamistingimuste parandamise kõrval oluline tegeleda ka aretusega, mis lühidalt tähendab soovitatavate isendite väljavalmist ja nende järglaskonna paljundamist.

Lihtsaim viis looma valikuks on võrrelda tema järglaste keskmisi näitajaid eakaaslaste keskmiste näitajatega. Sellist lihtsat meetodikat on loomakasvatajad kasutanud üle tuhande aasta ilma, et neil oleks aimugi olnud geneetikast ja statistikast. Tänapäeval on valiku põhimõte jäänud samaks, ainult et lihtsa võrdluse asemel püütakse arvesse võtta võimalikult palju mõjufaktoreid. Välja on töötatud rida meetodeid hindamiseks populatsiooni geneetilisi parameetreid, arvestades seejuures keskkonna võimalikku mõju.

Käesolevas töös vaadeldakse põhiliste aretuses kasutatavate geneetiliste parameetrite – dispersioonikomponentide ja nende kaudu avalduva päritavuskoefitsiendi – hindamist loomapopulatsioonides. Töö eesmärk on esitada algoritmid erinevate dispersioonikomponente hindavate meetodite jaoks, analüüsida saadavaid hinnanguid ja võrrelda hindamismeetodeid.

Töö koosneb viiest peatükist. Esimeses peatükis fikseeritakse geneetiline mudel, millest edasises töös lähtutakse, ning esitatakse vajalikud mõisted aretusteoorias. Teises peatükis fikseeritakse kasutatav statistiline mudel ning tuuakse kirjandusallikaist lähtudes sisse oluliseimad edaspidi kasutatavad mõisted, valemid ja omadused. Kolmas peatükk käsitleb dispersioonikomponentide ja päritavuskoefitsiendi hindamist tasakaalulise andmestiku korral; ühe juhusliku faktoriga mudeli korral tõestatakse kasutatavate ruutvormide jaotumine proportsionaalselt  $\chi^2$ -jaotusega, tuletatakse dispersioonikomponentide hinnangute dispersiooni valemid ja esitatakse päritavuskoefitsiendi hinnangu dispersiooni valem, tuletatakse tõenäosus, et päritavuskoefitsiendi hinnang ei

kuulu parameeterruumi ning viiakse paketi SAS abil läbi modelleerimiskatsed hinnangute omaduste uurimiseks ja teoreetiliste tulemuste kontrollimiseks. Neljandas peatükis vaadeldakse dispersioonikomponentide hindamist mittetasakaalulisel juhul; esitatakse viie meetodi algoritmid, neist MIVQUE, ML ja REML algoritmid on välja töötatud kirjanduses (Harville, 1977; Henderson, 1984; Searle jt., 1992; Mao, Schaeffer, 1993) avaldatud materjale kombineerides ning Hendersoni 3. meetod ja MIVQUE(0) meetod kirjandusest pärit vihjete ja eelduste alusel. Viiendas peatükis rakendatakse eelnevalt kirjeldatud meetodeid modelleeritud andmetele ja võrreldakse saadud dispersioonikomponentide ja päritavuskoefitsiendi hinnanguid.

# 1. Sissejuhatus aretusteooriasse

Põllumajanduslooma potentsiaalse jõudluse määrab ära tema genotüüp. Sellega annab geneetika teoreetilised alused loomade aretuseks ja selektsiooniks. Aretuse teel on võimalik parandada vaid looma neid omadusi, millel esineb pärilik muutlikkus. Seetõttu on oluline tunda tunnuste päritavuse astet ja geneetilise variatsiooni seaduspärasusi.

## 1.1. Lineaarne geneetiline mudel

Populatsiooni geneetiliste omaduste analüüsimiseks on esmalt vaja teada, millest on tingitud isendite fenotüübiväärtuste erinevus. On selge, et polümeerse tunnuse variatsioon populatsioonis on nii isendite genotüüpide kui ka neile mõjuvate keskkonnategurite erinevusest. Lihtsaim fenotüübiväärtust  $P$  kirjeldav mudel on lineaarne geneetiline mudel (Falconer, 1981):

$$P = G + E, \quad (1.1)$$

kus  $G$  on genotüübiväärtus (väljendab geneetilist osa tunnuse väärtusest) ja  $E$  väljendab keskkonna mõjusid.

Iga isendi geneetilist potentsiaali iseloomustab tema genotüübiväärtuse hälve populatsiooni keskmisest genotüübiväärtusest ehk genotüübiefekt (Teinberg, 1978):

$$g = G - \bar{G}. \quad (1.2)$$

Arvestades, et populatsiooni keskmine fenotüübiväärtus võrdub seda moodustavate isendite keskmise genotüübiväärtusega (keskmise keskkonna mõju võib lugeda võrdseks nulliga),  $\bar{P} = \bar{G}$ , on võrdus (1.2) kirjutatav ka kujul

$$g = G - \bar{P}.$$

Mudel (1.1) esitub siis järgmiselt:

$$P = \bar{P} + g + E.$$

Genotüübiefekt on, arvestades seda määravate alleelide vahelisi võimalikke seoseid, lahtikirjutatav summana:

$$g = A + D + I, \quad (1.3)$$

kus  $A$  on looma aretusväärtus ehk aditiivne geneetiline väärtus (s.o. genotüübis olevate alleelide summaarne efekt),  $D$  on alleelide lookustesisesest interaktsioonist ehk dominantsusest tingitud efekt (dominantsushälve) ja  $I$  on geenide mitteaditiivsest toimest

(erinevates lookustes asuvate geenipaaride koosmõjust) tingitud interaktsioonihälve ehk epistaasi efekt.

Et vanemalt järglastele päranduvad vaid üksikalleelid, mitte aga nendevahelised interaktsioonid, pakuvad aretuse seisukohast huvi ja on samas järglaste järgi hinnatavad üksnes alleeliväärtused. Keskmise alleeliväärtus (geeniväärtus) on hinnatav nende isendite hälbega populatsiooni keskmisest, kes said ühelt vanemalt selle alleeli, kusjuures kõik ülejäänud alleelid jaotuvad täiesti juhuslikult. Genotüüpi kuuluvate alleelide väärtuste summa ongi looma aretusväärtus  $A$ . Seega on summa (1.3) liidetavatest praktiliselt hinnatav vaid aretusväärtus, mistõttu praktiliste aretuslike probleemide lahendamisel loetakse genotüübiefekt  $g$  võrdseks aretusväärtusega  $A$  ning mudel (1.1) saab kuju (van Arendonk jt., 1997)

$$P = \bar{P} + A + E. \quad (1.4)$$

Siin eeldatakse, et aretusväärtuse keskmine üle populatsiooni on null:  $\bar{A} = 0$ . Aretusväärtuste alusel toimub põhiline selektsioonitöö ja valitakse parimad isendid populatsiooni taastootmiseks.

Vastavalt valemile (1.1) on isendite fenotüübiväärtuste dispersioon populatsioonis ehk lihtsalt fenotüübidispersioon  $\sigma_P^2$  genotüübi ja keskkonna sõltumatuse korral avaldatav genotüübidispersiooni  $\sigma_G^2$  ja keskkonnadispersiooni  $\sigma_E^2$  summana:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2.$$

Et aga sugulasisendite omavaheline sarnasus polümeerse tunnuse osas on tingitud peamiselt alleelide aditiivsest toimest, on praktikas enamkasutatav fenotüübidispersiooni komponentideks lahutus vastavalt valemile (1.4):

$$\sigma_P^2 = \sigma_A^2 + \sigma_E^2, \quad (1.5)$$

kus  $\sigma_A^2$  on alleelide aditiivse (vanemalt järglastele päranduva) toime dispersioon ehk aditiivdispersioon. Kui aditiivdispersioon võrdub nulliga, on loomadel identsed aretusväärtused.

Dispersioonikomponendid on hinnatavad sugulasrühmade dispersioonanalüüsi abil, kasutades valdavalt poolõvesid, harvem ka täisõvesid. Põllumajandusloomade korral jagatakse populatsioon tavaliselt rühmadeks isade alusel, sest üks ema võib oma elu jooksul ilmale tuua vaid loetud arvu järglasi, sama isa järglasi võib produtseerida aga

tuhandeid. Samuti elavad ühe ema järglased harilikult väga sarnastes tingimustes ja on tugevas seoses keskkonnaga, ühe isa järglased on aga hajutatud erinevate keskkonna- ja pidamistingimustega alale. Seega on ema efekt segatud keskkonna mõjuga.

Arvestades, et iga indiviidi aretusväärtusest  $A_i$  on pool tingitud isa mõjust  $\sigma_i$  (pooled alleelid on päritud isalt), s.t.  $\sigma_i = 1/2 A_i$ , ja teades, et aditiivdispersioon on defineeritud kui aretusväärtuse dispersioon, on isa efekti dispersioon avaldatav seosega

$$D(\sigma_i) = D(1/2 A_i) = 1/4 D(A_i),$$

s.t. aditiivdispersioon on võrdne neljakordse isast tingitud dispersioonikomponendiga,

$$\sigma_A^2 = 4\sigma_\sigma^2. \quad (1.6)$$

Analoogselt on aditiivdispersioon avaldatav emapoolsete poolõvede rühmade analüüsi korral,  $\sigma_A^2 = 4\sigma_\sigma^2$ , ja täisõvede rühmade analüüsi korral,  $\sigma_A^2 = 2(\sigma_\sigma^2 + \sigma_\varphi^2)$ .

## 1.2. Päritavuskoefitsient

Olulisim kvantitatiivseid tunnuseid iseloomustav geneetiline parameeter loomapopulatsioonides on päritavuskoefitsient ehk päritavus. Päritavus seob indiviidi omadused (fenotüübi) tema geneetilise konstitutsiooniga (genotüübiga).

Päritavus kitsamas mõttes, ehk lihtsalt päritavus, näitab, kui suur osa populatsiooni üldisest fenotüübilisest muutlikkusest on tingitud päritavast genotüübilisest muutlikkusest. Päritavus  $h^2$  avaldub genotüübi poolt põhjustatud päriliku muutlikkuse ehk aditiivdispersiooni  $\sigma_A^2$  ja fenotüübidispersiooni  $\sigma_P^2$  suhtena:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}. \quad (1.7)$$

Harvem leiab kasutamist päritavus laiemas mõttes ehk geneetilise determinatsiooni koefitsient  $h_w^2$ , mis näitab, mil määral erinevused isendite genotüüpides peegelduvad nende fenotüübiväärtuste erinevuses, ehk milline osa populatsiooni üldisest fenotüübilisest muutlikkusest on tingitud genotüübilisest muutlikkusest. Seega avaldub  $h_w^2$  genotüübidispersiooni  $\sigma_G^2$  ja fenotüübidispersiooni  $\sigma_P^2$  suhtena:

$$h_w^2 = \frac{\sigma_G^2}{\sigma_P^2}.$$

$$0 \leq \sigma_A^2 \leq \sigma_G^2 \leq \sigma_P^2,$$

siis

$$0 \leq h^2 \leq h_w^2 \leq 1.$$

Kui päritavuskoeffitsient võrdub nulliga, siis uuritava tunnuse hälbed populatsiooni keskmisest ei ole päritavad, ja kui päritavuskoeffitsient võrdub ühega, on kogu tunnuse muutlikkus seletatav aditiivse geneetilise mõjuga.

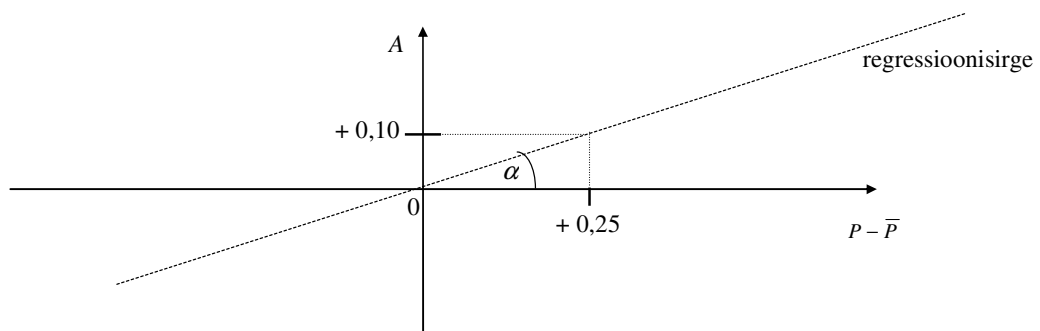
Päritavuskoeffitsiendi olulisus on tingitud sellest, et aretusväärtus on hinnatav kui indiviidi fenotüübiväärtuse hälve populatsiooni keskmisest fenotüübiväärtusest korrutatuna päritavuskoeffitsiendiga:

$$\hat{A} = h^2(P - \bar{P}). \quad (1.8)$$

Seos (1.8) järeldub vähimruutude printsiibi abil valemist (1.4). Sellest seosest nähtub, et päritavuskoeffitsient on defineeritav ka kui aretusväärtuse regressioon fenotüübiväärtusehälbele populatsiooni keskmise suhtes (Brascamp jt., 1995):

$$h^2 = b_{A/(P-\bar{P})}, \quad (1.9)$$

kus  $b_{A/(P-\bar{P})}$  on regressioonisirge tõus. Näiteks kui piima rasvaprotsendi päritavuskoeffitsient on  $h^2 = 0,40$  ja lehm annab keskmisest 0,25 % võrra rammusamat piima, siis tema aretusväärtuse hinnanguks on +0,10 (joonis 1.1). Kuna järglasele pärandub pool ema geenide aditiivsest mõjust, saab järglane emalt oodatavalt kaasa võime anda keskmisest 0,05 % võrra rammusamat piima.



**Joonis 1.1.** Regressioonisirge aretusväärtuse  $A$  prognoosimiseks fenotüübiväärtusehälbe  $P - \bar{P}$  abil,  $h^2 = \tan\alpha = 0,40$ .



Kuna aditiivdispersioon on isapoolsete poolõvede rühmade alusel täpsemini hinnatav kui emapoolsete poolõvede korral, hinnatakse päritavuskoeffitsienti isast tingitud dispersioonikomponendi abil valemeist (1.6) ja (1.7) järelduva seosega:

$$h^2 = \frac{4\sigma_\sigma^2}{\sigma_P^2}. \quad (1.10)$$

Päritavuskoeffitsiendi kasutamisel tuleb pöörata tähelepanu sellele, et  $h^2$  ei oma mingit absoluutset liigile omast väärtust, vaid on arvatud konkreetse populatsiooni jaoks konkreetsel ajahetkel. Näiteks tõumaterjali sissetoomine välismaalt suurendab ilmselt geneetilist varieeruvust ja seeläbi ka päritavuskoeffitsiendi väärtust. Varasemal ajahetkel või mõnes teises populatsioonis arvatud päritavuskoeffitsiendi kasutamine aretusprogrammi koostamisel võib viia valede prognoosideni, mille tagajärjed ilmnevad alles aastate pärast.

## 2. Üldine metoodika dispersioonikomponentide ja päritavuskoeffitsiendi hindamiseks

Põhiprobleem dispersioonikomponentide ja päritavuse ning seeläbi ka aretusväärtuse hindamisel fenotüübil mõõdetud tunnuste alusel on eraldada geneetilised mõjud keskkonna mõjudest. Matemaatiliselt tähendab see sobiva mudeli valikut ja selle parameetrite hindamist.

### 2.1. Üldine lineaarne mudel

Põllumajandusloomade aretuses kasutatavad statistilised meetodid on tuletatud üldisest lineaarsest mudelist või selle modifikatsioonidest lähtudes. Üldine lineaarne mudel (ÜLM) on uurija poolt eeldatav faktorite ja tunnuste vahekorra üldskeem, mille detailid tuleb täiendavalt hinnata algandmetest (valimist). ÜLM võimaldab analüüsida faktorite mõju uuritavatele tunnustele. Tehtavad järeldused võib lugeda õigeks, kui uuritav ÜLM vastab tegelikkusele ja kehtivad vajalikud matemaatilised eeldused (normaaljaotus, vaatluste sõltumatus jm.).

Üldine lineaarne mudel on maatrikskujul esitatav järgmiselt (Searle, 1987):

$$y = Xb + Zu + e, \quad (2.1)$$

ehk

$$y = Xb + Z_1u_1 + \dots + Z_su_s + e,$$

kus  $y$  on uuritava tunnuse juhuslik  $N \times 1$  vektor,

$b$  on tundmatu fikseeritud efektide  $p \times 1$  vektor,

$u$  on tundmatu juhuslike efektide  $n \times 1$  vektor,

$X$  on fikseeritud efektidele vastav  $N \times p$  plaanimaatriks,

$Z$  on juhuslikele efektidele vastav  $N \times n$  plaanimaatriks ( $Z_i : N \times n_i$ ),

$e$  on juhuslike vigade  $N \times 1$  vektor.

Juhuslike muutujate keskvärtused fikseeritakse järgmiselt:

$$E(u)=0, \quad E(e)=0, \quad E(y)=Xb.$$

Kui dispersioonimaatriksid on defineeritud kujul

$$G = D(u) = D \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_s \end{bmatrix} = \begin{bmatrix} I_{n_1} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & I_{n_2} \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_s} \sigma_s^2 \end{bmatrix}, \quad (2.2)$$

kus  $n_i$  on juhusliku faktori  $u_i$  tasemete arv,  $\sum_{i=1}^s n_i = n$ ,

$$R = D(e) = I_N \sigma_e^2, \quad (2.3)$$

$$\text{Cov}(u_i, u_j) = 0, \quad i \neq j,$$

ja

$$\text{Cov}(u, e') = 0,$$

siis

$$V = D(y) = ZGZ' + R = \sum_{i=1}^s Z_i Z_i' \sigma_i^2 + I_N \sigma_e^2. \quad (2.4)$$

Tähistades  $V_0 = I$ ,  $\sigma_0^2 = \sigma_e^2$  ja  $V_i = Z_i Z_i'$  saame

$$V = \sum_{i=0}^s Z_i Z_i' \sigma_i^2 = \sum_{i=0}^s V_i \sigma_i^2. \quad (2.5)$$

Vähimruutude hinnangud parameetritele  $b$  ja  $u$  on leitavad normaalvõrrandist

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}. \quad (2.6)$$

Juhuslike efektide mõju on prognoositav Hendersoni segamudeli võrrandist

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}, \quad (2.7)$$

mis on, arvestades maatriksi  $R$  esitust (2.3), kirjutatav ka kujul

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \sigma_e^2 \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}. \quad (2.8)$$

Siit tuleneb, et juhuslike efektide vektori  $u$  parim lineaarne nihketa prognoos on

$$\hat{u} = (Z'Z + G^{-1} \sigma_e^2)^{-1} Z'(y - X\tilde{b}), \quad (2.9)$$

kus  $\tilde{b} = (X'V^{-1}X)^{-1} X'V^{-1}y$  on fikseeritud efektide üldine vähimruutude hinnang (Henderson, 1984). Võrdus (2.9) on, arvestades üldist seost

$$(Z'R^{-1}Z + G^{-1})^{-1} Z'R^{-1} = GZ'V^{-1}$$

(Searle jt., 1992), esitatav kujul

$$\hat{u} = GZ'V^{-1}(y - X\tilde{b}). \quad (2.10)$$

Aretusteoorias lihtsustatakse mudelit, liites kõigi fikseeritud faktorite mõjud üheks, nn. karja-aasta-sesooni efektiks. Juhusliku faktorina käsitletakse isa, millest tuleneb, et juhuslike efektide vektor  $u$  on isade aretusväärtuste vektor. Aretuses on võrdus (2.10) tuntud rohkem ühemõõtmelisel juhul:

$$\hat{u}_i = \frac{n_i \sigma_\sigma^2}{n_i \sigma_\sigma^2 + \sigma_e^2} (\bar{y}_i - \mu),$$

mis, kasutades päritavuskoeffitsiendi arvutamise valemit (1.10), on ümber kirjutatav kujul

$$\hat{u}_i = \frac{n_i}{n_i + (4/h^2 - 1)} (\bar{y}_i - \mu) = \frac{n_i h^2}{(n_i - 1)h^2 + 4} (\bar{y}_i - \mu).$$

Siin  $\hat{u}_i$  on isa  $i$  aretusväärtuse prognoos,  $n_i$  on tütarde arv isal  $i$ ,  $\bar{y}_i$  on uuritava tunnuse keskmine isa  $i$  järglastel ja  $\mu$  on üldkeskmine (populatsiooni keskmine) (Searle jt., 1992).

## 2.2. Dispersioonikomponentide hindamine üldisest lineaarsest mudelist

Parim mudel esitamaks erinevaid dispersioonikomponentide hindamise meetodeid on ÜLM (2.1). Kõik meetodid põhinevad normaalvõrrandi (2.6) või segamudeli võrrandi (2.7) parema poole ruutvormide kasutamisel.

Ruutvorm on reaalsete kordajatega homogeenne ruutpolünoom  $y'Qy$ , kus  $Q$  on sümmeetriline maatriks. Kui  $E(y) = Xb$  ja  $D(y) = V$ , siis iseloomustavad ruutvormi järgmised omadused (Searle jt., 1992):

1.  $E(y'Qy) = tr(QV) + b'X'QXb$ ;
2.  $D(y'Qy) = 2tr(QVQV) + 4b'X'QXb$ , kui  $y \sim N(Xb, V)$ ;
3. kui  $y \sim N(Xb, V)$  ja  $QVQV = QV$  ( $QV$  on idempotentne), siis

$$y'Qy \sim \chi^2(r(Q), \frac{1}{2}b'X'QXb);$$

4. kui  $y \sim N(Xb, V)$ , siis on ruutvormid  $y'Qy$  ja  $y'Py$  sõltumatud tingimusel

$$QVP = 0.$$

Enamus dispersioonikomponente hindavaid meetodeid kasutab ühtset põhistrateegiat (Mao, Schaeffer, 1993):

1. defineeritakse vajalikud ruutvormid (keskruudud);
2. võrdsustatakse ruutvormide arvvaartused nende oodatavate väärtustega vastavalt meetodi reeglitele; oodatavad väärtused on siis tundmatute dispersioonikomponentide funktsioonid;
3. avaldatakse tundmatud komponendid, kasutades maatriksalgebra vahendeid või iteratsiooniprotsessi.

On kolm printsiipi, mida dispersioonikomponentide hindamiseks kasutatavate ruutvormide  $y'Qy$  korral tavaliselt kontrollitakse (Rao, 1972; Searle jt., 1992).

1) Ruutvormide invariantisus parameetri  $b$  suhtes. Ruutvorm  $y'Qy$  on invariantne, kui, vaadates suuruse  $b$  asemel suurust  $b+k$  mudelis (2.1), kehtib võrdus

$$y'Qy = (y + Xk)'Q(y + Xk)$$

iga vektori  $k$  korral. Siit

$$y'Qy = y'Qy + 2y'QXk + k'X'QXk ,$$

millest järeldub, et

$$QX = 0 .$$

Ehk, arvestades ruutvormi keskvaartuse avaldist,

$$E(y'Qy) = tr(QV) .$$

Seega ei mõjuta fikseeritud efektid invariantsete ruutvormide keskvaartust. Invariantisus on ainus omadus, mis kehtib peaaegu kõigi dispersioonikomponente hindavate meetodite korral.

2) Hinnangute nihketus. Selleks vaatame üksiku dispersioonikomponendi  $\sigma_i^2$  hindamise asemel dispersioonikomponentide lineaarkombinatsiooni  $p'\sigma^2$  hindamist ja nõuame, et hindamiseks kasutatava ruutvormi  $y'Qy$  ooteväärtus võrduks hinnatava suurusega  $p'\sigma^2$ :

$$E(y'Qy) = b'X'QXb + \sum_{i=0}^r tr(QZ_iZ_i')\sigma_i^2 = p'\sigma^2 .$$

Invariantse hinnangu korral  $b'X'QXb = 0$  ning nihketa hinnangu saamiseks on vajalik, et

$$tr(QZ_iZ_i') = p_i .$$

3) Hinnangute minimaalne varieeruvus, mis invariantsete ja nihketa hinnangute ning normaaljaotuse korral tähendab suuruse

$$D(y'Qy) = 2tr(QVQV) + 4b'X'QXb = 2tr(QVQV)$$

minimiseerimist.

### 3. Dispersioonikomponentide ja päritavuskoefitsiendi hindamine tasakaalulisel juhul

#### 3.1. ANOVA hinnangud

Tasakaalulise andmestiku korral toimub dispersioonikomponentide hindamine otse dispersioonanalüüsi tabelist. Selleks võrdsustatakse leitud keskrüüud ehk hälvete ruutude summa abil arvutatud dispersiooni hinnangud vastava hinnangu keskvaertusega. Saadud võrranditest avaldatakse soovitud dispersioonikomponendid.

Vaatame ühe juhusliku faktoriga mudelit

$$y_{ij} = \mu + \sigma_i + e_{ij}, \quad (3.1)$$

kus  $y_{ij}$  on tunnuse väärtus isa  $i$  järglasel  $j$ ,  $i = 1, \dots, a$  ja  $j = 1, \dots, n$ . Oletame siin, et isa on valitud juhuslikult ja seega on isa mõju  $\sigma_i$  juhuslik. Vabaliige  $\mu$  on üldkeskmine ja  $e_{ij}$  on isa valikust sõltumatu juhuslik viga.

Esitusest (3.1) leiame fenotüübidispersiooni  $\sigma_p^2$ :

$$D(y_{ij}) = \sigma_p^2 = \sigma_\sigma^2 + \sigma_e^2.$$

Suuruste  $\sigma_\sigma^2$  ja  $\sigma_e^2$  hindamiseks lähtume tabelist 3.1, võttes seal  $MSA = n\hat{\sigma}_\sigma^2 + \hat{\sigma}_e^2$  ja  $MSE = \hat{\sigma}_e^2$ . Saadud hinnangud

$$\hat{\sigma}_\sigma^2 = \frac{MSA - MSE}{n} \quad (3.2)$$

ja

$$\hat{\sigma}_e^2 = MSE \quad (3.3)$$

on nihketa, sest  $E(\hat{\sigma}_e^2) = E(MSE) = \sigma_e^2$  ja

$$E(\hat{\sigma}_\sigma^2) = \frac{E(MSA) - E(MSE)}{n} = \frac{n\sigma_\sigma^2 + \sigma_e^2 - \sigma_e^2}{n} = \sigma_\sigma^2.$$

Kui  $\hat{\sigma}_\sigma^2$  on isast tingitud dispersioonikomponendi hinnang ja  $n$  on tütarde arv ühel isal, siis saame valemist (1.10) päritavuskoefitsiendi hinnangu

$$\hat{h}^2 = \frac{4\hat{\sigma}_\sigma^2}{\hat{\sigma}_\sigma^2 + \hat{\sigma}_e^2} = \frac{4(MSA - MSE)}{MSA + (n-1)MSE}. \quad (3.4)$$

**Tabel 3.1.** Dispersioonanalüüsi tabel ühe juhusliku faktori korral tasakaalulisel juhul (Tiit, Parring, Möls, 1977).

Hajuvuse allikas	Vabadusastmete arv	Ruutvorm	Dispersiooni hinnang	Hinnangu keskväärts
Isa	$a-1$	$SSA = \sum_i n(\bar{y}_i - \bar{y}_{..})^2$	$MSA = SSA / (a - 1)$	$E(MSA) = n\sigma_{\%}^2 + \sigma_e^2$
Jääk	$a(n-1)$	$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$	$MSE = SSE / a(n-1)$	$E(MSE) = MSE = \sigma_e^2$
Kokku	$an-1$	$SST = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$		

Esitamaks mudelit (3.1) maatrikskujul ja täpsustamaks tehtavaid eeldusi, võtame kasutusele ühtedest koosneva  $(n \times 1)$ -vektori  $1_n$  ja  $(n \times n)$ -maatriksid  $J_n = 1_n 1_n'$  ning  $\bar{J}_n = \frac{1}{n} J_n$ . Siis saab võrdus (3.1) kuju

$$y = (1_a \otimes 1_n) \mu + (I_a \otimes 1_n) \sigma + e.$$

Edaspidi eeldame mitmemõõtmelist normaaljaotust

$$\begin{bmatrix} \sigma \\ e \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_a \sigma_{\sigma}^2 & 0 \\ 0 & I_{an} \sigma_e^2 \end{bmatrix} \right),$$

$$y \sim N(1_{an} \mu, V), \quad V = D(y) = I_a \otimes (\sigma_{\sigma}^2 J_n + \sigma_e^2 I_n).$$

Kasutatavad ruutvormid maatrikskujul on järgmised:

$$SSA = y' [(I_a - \bar{J}_a) \otimes \bar{J}_n] y = y' A_1 y,$$

$$SSE = y' [I_{an} - (I_a \otimes \bar{J}_n)] y = y' A_2 y.$$

Näitame, et

$$SSA / (n\sigma_{\sigma}^2 + \sigma_e^2) \sim \chi_{a-1}^2 \quad (3.5)$$

ja

$$SSE / \sigma_e^2 \sim \chi_{a(n-1)}^2. \quad (3.6)$$

Tõestuseks kasutame punktis 2.2 toodud ruutvormide kolmandat omadust. Tõestame esmalt seose (3.5).

Märgime, et maatriks  $\bar{J}_n = \frac{1}{n} 1_n 1_n'$  on idempotentne, sest

$$(\bar{J}_n)^2 = \frac{1}{n^2} 1_n \underbrace{1_n' \cdot 1_n}_{n} 1_n' = \frac{1}{n} 1_n 1_n' = \bar{J}_n.$$

Kuna

$$A_1 V = [(I_a - \bar{J}_a) \otimes \bar{J}_n] [I_a \otimes (\sigma_{\sigma}^2 J_n + \sigma_e^2 I_n)] =$$



$$\begin{aligned}
&= (I_a - \bar{J}_a) \otimes \bar{J}_n (\sigma_\sigma^2 \bar{J}_n n + \sigma_e^2 I_n) = \\
&= (I_a - \bar{J}_a) \otimes \bar{J}_n (n\sigma_\sigma^2 + \sigma_e^2) = \\
&= (n\sigma_\sigma^2 + \sigma_e^2) [(I_a - \bar{J}_a) \otimes \bar{J}_n] = (n\sigma_\sigma^2 + \sigma_e^2) A_1,
\end{aligned}$$

siis on maatriks  $\frac{A_1 V}{n\sigma_\sigma^2 + \sigma_e^2}$  idempotentne, sest

$$\begin{aligned}
\left( \frac{A_1 V}{n\sigma_\sigma^2 + \sigma_e^2} \right) \left( \frac{A_1 V}{n\sigma_\sigma^2 + \sigma_e^2} \right) &= A_1 A_1 = [(I_a - \bar{J}_a) \otimes \bar{J}_n] [(I_a - \bar{J}_a) \otimes \bar{J}_n] = \\
&= (I_a - \bar{J}_a)^2 \otimes \bar{J}_n^2 = (I_a - \bar{J}_a) \otimes \bar{J}_n = A_1 = \frac{A_1 V}{n\sigma_\sigma^2 + \sigma_e^2}.
\end{aligned}$$

Jääb veel märkida, et

$$r(A_1) = \text{tr}(A_1) = a - 1.$$

Tõestuse lõpetamiseks näitame, et  $\mu 1'_{an} A_1 = 0$ :

$$1'_{an} A_1 = (1'_a \otimes 1'_n) [(I_a - \bar{J}_a) \otimes \bar{J}_n] = (1'_a - \underbrace{1'_a \cdot \frac{1}{a} 1_a 1'_a}_1) \otimes (1'_n \cdot \bar{J}_n) = 0.$$

Seose (3.6) tõestuseks saame analoogselt, et kuna

$$\begin{aligned}
A_2 V &= [I_{an} - (I_a \otimes \bar{J}_n)] [I_a \otimes (\sigma_\sigma^2 J_n + \sigma_e^2 I_n)] = \\
&= \sigma_\sigma^2 (I_a \otimes J_n - I_a \otimes \underbrace{J_n \bar{J}_n}_{\bar{J}_n}) + \sigma_e^2 (I_{an} - I_a \otimes \bar{J}_n) = \\
&= \sigma_\sigma^2 \cdot 0 + \sigma_e^2 (I_{an} - I_a \otimes \bar{J}_n) = \sigma_e^2 A_2,
\end{aligned}$$

siis on maatriks  $\frac{A_2 V}{\sigma_e^2}$  idempotentne, sest

$$\begin{aligned}
\left( \frac{A_2 V}{\sigma_e^2} \right) \left( \frac{A_2 V}{\sigma_e^2} \right) &= A_2 A_2 = [I_{an} - (I_a \otimes \bar{J}_n)] [I_{an} - (I_a \otimes \bar{J}_n)] = \\
&= I_{an} - 2(I_a \otimes \bar{J}_n) + I_a \otimes \bar{J}_n = I_{an} - (I_a \otimes \bar{J}_n) = A_2 = \frac{A_2 V}{\sigma_e^2}.
\end{aligned}$$

Tõestuse lõpetuseks märgime, et

$$r(A_2) = \text{tr}(A_2) = a(n - 1)$$

ja

$$\mu 1'_{an} A_2 = \mu (1'_a \otimes 1'_n) [I_{an} - (I_a \otimes \bar{J}_n)] = 0.$$

Näitame veel, et ruutvormid  $SSA$  ja  $SSE$  on sõltumatud, kasutades punktis 2.2 toodud ruutvormide neljandat omadust. Saame

$$\begin{aligned} A_1 V A_2 &= [(I_a - \bar{J}_a) \otimes \bar{J}_n] I_a \otimes (\sigma_\sigma^2 J_n + \sigma_e^2 I_n) [I_{an} - (I_a \otimes \bar{J}_n)] = \\ &= (n\sigma_\sigma^2 + \sigma_e^2) [(I_a - \bar{J}_a) \otimes \bar{J}_n] [I_{an} - (I_a \otimes \bar{J}_n)] = \\ &= (n\sigma_\sigma^2 + \sigma_e^2) \{[(I_a - \bar{J}_a) \otimes \bar{J}_n] - [(I_a - \bar{J}_a) \otimes \bar{J}_n]\} = \\ &= (n\sigma_\sigma^2 + \sigma_e^2) \cdot 0 = 0. \end{aligned}$$

### 3.2. Dispersioonikomponentide ja päritavuskoeffitsiendi hinnangute varieeruvus

Ruutvormide  $SSA$  ja  $SSE$  sõltumatuse ning proportsionaalselt  $\chi^2$ -jaotusega jaotumise põhjal on lihtne tuletada dispersioonikomponentide hinnangute dispersiooni hindamise valemeid. Tabelist 3.1 ja valemist 3.6 järeldeb, et

$$SSE/E(MSE) \sim \chi_{a(n-1)}^2.$$

Teades, et  $\chi^2$ -jaotusega juhusliku suuruse dispersioon võrdub kahekordse vabadusastmete arvuga, saame

$$D(SSE/E(MSE)) = 2a(n-1).$$

Seega

$$D(SSE) = 2a(n-1)[E(MSE)]^2$$

ning

$$D(MSE) = D\left(\frac{SSE}{a(n-1)}\right) = \frac{2[E(MSE)]^2}{a(n-1)}. \quad (3.7)$$

Arvestades, et  $E(MSE) = \sigma_e^2$  ja  $MSE = \hat{\sigma}_e^2$ , saame dispersioonikomponendi hinnangu  $\hat{\sigma}_e^2$  dispersiooni kujul

$$D(\hat{\sigma}_e^2) = D(MSE) = \frac{2[E(MSE)]^2}{a(n-1)} = \frac{2\sigma_e^4}{a(n-1)}. \quad (3.8)$$

Leiame valemi (3.8) parema poole nihketa hinnangu. Võrdsustades avaldise (3.7) avaldisega

$$D(MSE) = E[MSE - E(MSE)]^2 = E[(MSE)^2] - [E(MSE)]^2,$$

saame seose

$$E[(MSE)^2] = [E(MSE)]^2 \left(1 + \frac{2}{a(n-1)}\right),$$

millest

$$\frac{E[(MSE)^2]}{a(n-1)+2} = \frac{[E(MSE)]^2}{a(n-1)}. \quad (3.9)$$

Seega on  $\frac{\hat{\sigma}_e^4}{a(n-1)+2}$  suuruse  $\frac{\sigma_e^4}{a(n-1)}$  nihketa hinnanguks. Esitusest (3.8) saame

nüüd  $\hat{\sigma}_e^2$  dispersiooni nihketa hinnangu valemi

$$\hat{D}(\hat{\sigma}_e^2) = \frac{2\hat{\sigma}_e^4}{a(n-1)+2}. \quad (3.10)$$

Analoogiliselt valemitele (3.7) ja (3.9) on saadav seosed

$$D(MSA) = \frac{2[E(MSA)]^2}{a-1} \quad (3.11)$$

ning

$$\frac{E[(MSA)^2]}{(a-1)+2} = \frac{[E(MSA)]^2}{a-1}. \quad (3.12)$$

Dispersioonikomponendi  $\sigma_\sigma^2$  hinnangu dispersioon avaldub valemite (3.8) ja (3.11) põhjal:

$$\begin{aligned} D(\hat{\sigma}_\sigma^2) &= D\left(\frac{MSA - MSE}{n}\right) = \frac{2}{n^2} \left\{ \frac{[E(MSA)]^2}{a-1} + \frac{[E(MSE)]^2}{a(n-1)} \right\} = \\ &= \frac{2}{n^2} \left[ \frac{(n\sigma_\sigma^2 + \sigma_e^2)^2}{a-1} + \frac{\sigma_e^4}{a(n-1)} \right]. \end{aligned} \quad (3.13)$$

Võttes seoses (3.12)  $E[(MSA)^2] = (n\hat{\sigma}_\sigma^2 + \hat{\sigma}_e^2)^2$  ja seoses (3.9)  $E[(MSE)^2] = \hat{\sigma}_e^4$ , on dispersioon (3.13) nende seoste alusel nihketa hinnatav valemist

$$\hat{D}(\hat{\sigma}_\sigma^2) = \frac{2}{n^2} \left[ \frac{(n\hat{\sigma}_\sigma^2 + \hat{\sigma}_e^2)^2}{a+1} + \frac{\hat{\sigma}_e^4}{a(n-1)+2} \right]. \quad (3.14)$$

Päritavuskoeffitsiendi hinnangu varieeruvus on poolõvede analüüsi korral hinnatav valemist (Becker, 1984)

$$D(\hat{h}^2) \approx 16 \frac{2[1 + (n-1)\rho]^2 (1-\rho)^2}{n(n-1)(a-1)}, \quad (3.15)$$

kus

$$\rho = \frac{\sigma_{\sigma}^2}{\sigma_{\sigma}^2 + \sigma_e^2} = \frac{1}{4}h^2 \quad (3.16)$$

on intraklasskorrelatsioonikordaja.

Saadud tulemuste kontrollimiseks viidi läbi modelleerimiskatsed, genereerides juhuslikud tasakaalulised andmestikud dispersioonikomponentide suhte

$$\tau = \sigma_{\sigma}^2 / \sigma_e^2 \quad (3.17)$$

väärtustel 0,01; 0,05; 0,1 ja 0,25. Selleks genereeriti isa mõju standardse normaaljaotusega ja juhuslik viga normaaljaotustega  $N(0, 100)$ ,  $N(0, 20)$ ,  $N(0, 10)$  ning  $N(0, 4)$ . Kuna

$$h^2 = 4\tau / (1 + \tau), \quad (3.18)$$

siis vastavad päritavuskoefitsiendi väärtused on 0,0396; 0,1905; 0,3636 ja 0,8. Isade arv  $a$  ning tütarde arv  $n$  valiti väärtuspaaridena  $(a, n) = (5, 5)$ ,  $(10, 25)$  ja  $(25, 100)$ . Genereeritud andmetega teostati protseduuri ANOVA abil dispersioonanalüüs ja arvutati valemite (3.2), (3.3) ja (3.4) dispersioonikomponentide ja päritavuskoefitsiendi hinnangud. Kogu protsessi korrati SAS-makro abil 1000 korda. Modelleerimise tulemused on toodud tabelis 3.2. Samas tabelis on ka valemite (3.8), (3.13) ja (3.15) abil leitud parameetrite hinnangute standardhälbed.

Nagu tabelist näha, on hinnangute keskmised ja nende standardhälbed üsna lähedased teoreetilistele väärtustele. Päritavuskoefitsiendi hinnangu keskvärtus on 11 juhul 12-st veidi väiksem tegelikust väärtusest. Ilmselt alahindab valem (3.4)  $h^2$  väärtust. Sama hakkab silma ka jooniselt 3.3, kus kõigi kolme modelleerimise korral on tõenäosim päritavuskoefitsiendi hinnang tegelikust väärtusest väiksem (vt. ka Kaart, 1995; 1996). Samuti on päritavuskoefitsiendi hinnangu varieeruvus pisut väiksem valemi (3.15) järgi arvatust. Võib arvata, et valem (3.15) hindab päritavuskoefitsiendi varieeruvust üle.

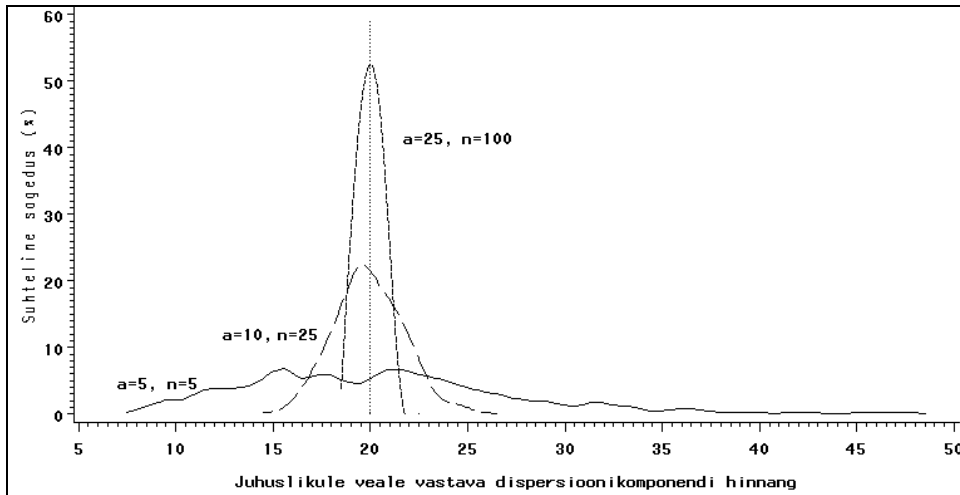
Mida suurem on isade ja tütarde arv, seda täpsem on nii dispersioonikomponentide kui ka päritavuskoefitsiendi hinnang (vt. ka joonised 3.1, 3.2 ja 3.3), ning seda lähedasemad on valemi (3.15) järgi leitud ja modelleerimise tulemusel saadud päritavuskoefitsiendi hinnangu standardhälbed. Samuti hakkab silma, et dispersioonikomponentide suhte  $\tau$  (ja seega ka  $h^2$ ) suurenemisel hinnangute varieeruvus väheneb.

**Tabel 3.2.** Parameetrite  $\sigma_e^2$ ,  $\sigma_\sigma^2$  ja  $h^2$  hinnangute modelleerimise tulemused.

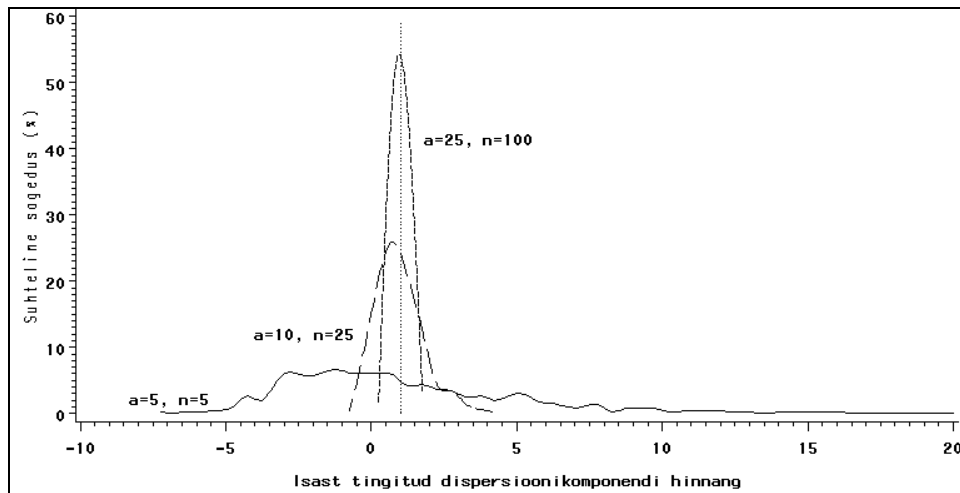
Teoreetilised standardhälbed on leitud valemitest (3.8), (3.13) ja (3.15).

Isade arv ( $a$ ) ja tütarde arv isal ( $n$ )	Parameetri tegelik väärtus	Hinnangu teoreetiline st. hälve	Hinnangu empiirilise jaotuse karakteristikud		
			Keskmine	St. hälve	Asümmeetria- kordaja
$a = 5, n = 5$	$\sigma_e^2 = 100$	31,6228	101,5827	33,3351	0,8516
	$\sigma_\sigma^2 = 1$	13,6748	-0,2669	15,8890	1,0010
	$h^2 = 0,0396$	0,6510	-0,0303	0,5868	0,8397
$a = 10, n = 25$	$\sigma_e^2 = 100$	9,1287	99,9983	8,8625	0,1447
	$\sigma_\sigma^2 = 1$	2,1631	0,9030	2,2338	0,8144
	$h^2 = 0,0396$	0,0943	0,0352	0,0872	0,7112
$a = 25, n = 100$	$\sigma_e^2 = 100$	2,8427	99,9041	2,8367	0,0825
	$\sigma_\sigma^2 = 1$	0,5554	1,0180	0,5750	0,6468
	$h^2 = 0,0396$	0,0228	0,0403	0,0226	0,6101
$a = 5, n = 5$	$\sigma_e^2 = 20$	6,3246	20,2330	6,8370	0,6747
	$\sigma_\sigma^2 = 1$	3,1517	0,8756	3,8848	1,1394
	$h^2 = 0,1905$	0,7171	0,1385	0,6701	0,6689
$a = 10, n = 25$	$\sigma_e^2 = 20$	1,8257	19,9305	1,8272	0,1915
	$\sigma_\sigma^2 = 1$	0,7710	0,9611	0,8141	0,7514
	$h^2 = 0,1905$	0,1571	0,1813	0,1496	0,6358
$a = 25, n = 100$	$\sigma_e^2 = 20$	0,5685	20,0153	0,5947	0,1133
	$\sigma_\sigma^2 = 1$	0,3329	1,0011	0,2600	0,1933
	$h^2 = 0,1905$	0,0632	0,1902	0,0478	0,1364
$a = 5, n = 5$	$\sigma_e^2 = 10$	3,1623	10,1857	3,1540	0,4905
	$\sigma_\sigma^2 = 1$	1,8439	0,9821	2,1238	1,1106
	$h^2 = 0,3636$	0,7840	0,3068	0,6705	0,4944
$a = 10, n = 25$	$\sigma_e^2 = 10$	0,9129	9,9863	0,9140	0,2509
	$\sigma_\sigma^2 = 1$	0,5981	0,9852	0,6423	0,8409
	$h^2 = 0,3636$	0,2227	0,3509	0,2121	0,5310
$a = 25, n = 100$	$\sigma_e^2 = 10$	0,2843	10,0017	0,2869	0,0083
	$\sigma_\sigma^2 = 1$	0,3051	1,0050	0,3153	0,4941
	$h^2 = 0,3636$	0,1055	0,3625	0,1033	0,3399
$a = 5, n = 5$	$\sigma_e^2 = 4$	1,2649	4,1196	1,2264	0,4201
	$\sigma_\sigma^2 = 1$	1,0696	0,9277	1,2831	1,6325
	$h^2 = 0,8$	0,9107	0,6225	0,7663	0,2574
$a = 10, n = 25$	$\sigma_e^2 = 4$	0,3652	3,9952	0,3754	0,3061
	$\sigma_\sigma^2 = 1$	0,4948	0,9951	0,5333	0,8004
	$h^2 = 0,8$	0,3572	0,7667	0,3336	0,2841
$a = 25, n = 100$	$\sigma_e^2 = 4$	0,1137	3,9983	0,1169	0,0970
	$\sigma_\sigma^2 = 1$	0,2885	1,0005	0,1624	0,2248
	$h^2 = 0,8$	0,1931	0,7977	0,1058	-0,0049

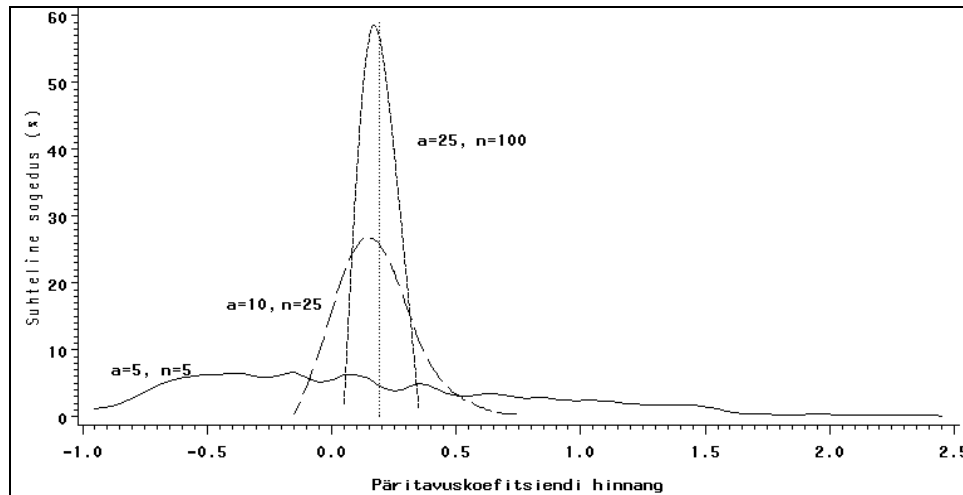
Joonistelt 3.1, 3.2 ja 3.3 on selgelt näha väärtuste paiknemise ebasümmeetrilise - keskväärtusest suuremad väärtused tekitavad väljavenitatud parempoolse „saba”. Sama kinnitavad ka asümmeetriakordaja positiivsed väärtused tabelis 3.2. Päritavauskoeffitsiendi väärtuse suurenedes hinnangute ebasümmeetrilisus väheneb, olles 25 isa, 25×100 tütre ja  $h^2 = 0,8$  korral peaaegu null.



**Joonis 3.1.** Juhuslikule veale vastava dispersioonikomponendi hinnangu  $\hat{\sigma}_e^2$  empiiriline jaotus isade arvu  $a$  ja tütarde arvu  $n$  väärtustepaaride  $(a, n) = (5, 5), (10, 25), (25, 100)$  ning  $\sigma_e^2 = 20$  korral.



**Joonis 3.2.** Isast tingitud dispersioonikomponendi hinnangu  $\hat{\sigma}_\sigma^2$  empiiriline jaotus isade arvu  $a$  ja tütarde arvu  $n$  väärtustepaaride  $(a, n) = (5, 5), (10, 25), (25, 100)$  ning  $\sigma_\sigma^2 = 1$  ja  $\sigma_e^2 = 20$  korral.



**Joonis 3.3.** Päritavuskoeffitsiendi hinnangu  $\hat{h}^2$  empiiriline jaotus isade arvu  $a$  ja tütarde arvu  $n$  väärtustepaaride  $(a, n) = (5, 5), (10, 25), (25, 100)$  ning  $h^2 = 0,1905$  korral.

### 3.3. Päritavuskoeffitsiendi illegaalse hinnangu saamise tõenäosus

Definitsiooni kohaselt on dispersioonikomponendid  $\sigma_\sigma^2$  ja  $\sigma_p^2$  alati positiivsed ning  $0 \leq 4\sigma_\sigma^2 \leq \sigma_p^2$ . Seetõttu on päritavuskoeffitsiendi väärtus alati nulli ja ühe vahel. Dispersioonikomponentide hinnangud võivad aga vahel ka antud piirkonnast välja jääda ja siis tuleb päritavuskoeffitsiendi hinnang illegaalne, s.t. negatiivne või suurem kui 1 (vt. joonised 3.2 ja 3.3). Tasakaalulise ühefaktorilise mudeli korral on võimalik leida tõenäosus, et selline olukord tekib.

Järgnevalt leiamegi  $h^2$  illegaalse hinnangu tõenäosuse  $P_I$  tasakaalulise ühefaktorilise mudeli korral. Selleks toome sisse tähistuse  $F$ , mida kasutame edaspidi kahes tähenduses. Esiteks, kirjutades  $X \sim F_{k,l}$ , tähendab see, et juhuslik suurus  $X$  on  $F$ -jaotusega vabadusastmetega  $k$  ja  $l$ . Teiseks, kirjutades  $P\{F_{k,l} > x_\alpha\} = \alpha$ , tähistab  $F_{k,l}$  juhuslikku suurust, mis on  $F$ -jaotusega vabadusastmetega  $k$  ja  $l$ .

Sündmused, et päritavuskoeffitsiendi hinnang on väiksem nullist või suurem ühest, on üksteist välistavad ja seega avaldub nende sündmuste summa tõenäosus liidetavate tõenäosuste summana:

$$P_I = P\{\hat{h}^2 < 0 \cup \hat{h}^2 > 1\} = P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\}.$$

Päritavuskoeffitsiendi valemist (3.4) järeldub, et

$$\hat{h}^2 < 0, \text{ kui } -\hat{\sigma}_e^2 < \hat{\sigma}_\sigma^2 < 0 \quad (\hat{\sigma}_e^2 \text{ on alati positiivne)}$$

ja

$$\hat{h}^2 > 1, \text{ kui } \hat{\sigma}_\sigma^2 / \hat{\sigma}_e^2 > 1/3 \text{ või } \hat{\sigma}_\sigma^2 < -\hat{\sigma}_e^2.$$

Tuletame esmalt päritavuskoeffitsiendi negatiivse hinnangu tõenäosuse:

$$P\{\hat{h}^2 < 0\} = P\{-\hat{\sigma}_e^2 < \hat{\sigma}_\sigma^2 < 0\} = P\{\hat{\sigma}_\sigma^2 < 0\} - P\{\hat{\sigma}_\sigma^2 < -\hat{\sigma}_e^2\}.$$

Valemite (3.2) ja (3.3) alusel saame, et

$$P\{\hat{\sigma}_\sigma^2 < -\hat{\sigma}_e^2\} = P\{MSA < (1-n)MSE\} = 0$$

ning seega

$$P\{\hat{h}^2 < 0\} = P\{\hat{\sigma}_\sigma^2 < 0\} = P\{MSA < MSE\} = P\left\{\frac{MSA}{MSE} < 1\right\}.$$

Lähtudes valemist (3.5) ja (3.6) ning sellest, et  $MSA$  ja  $MSE$  on tasakaalulise andmestiku korral sõltumatud, saame, et

$$\frac{MSA}{n\sigma_\sigma^2 + \sigma_e^2} \bigg/ \frac{MSE}{\sigma_e^2} \sim F_{(a-1), a(n-1)}.$$

Tähistades  $F = MSA / MSE$ , võime jätkata:

$$P\{F < 1\} = P\left\{\left(\frac{n\sigma_\sigma^2}{\sigma_e^2} + 1\right) F_{(a-1), a(n-1)} < 1\right\} = P\left\{F_{(a-1), a(n-1)} < \frac{\sigma_e^2}{\sigma_e^2 + n\sigma_\sigma^2}\right\}.$$

Tuues sisse valemiga (3.17) defineeritud dispersioonikomponentide suhte  $\tau$  ja valemiga (3.16) defineeritud intraklasskorrelatsioonikordaja  $\rho$ , saab negatiivse  $\hat{h}^2$  tõenäosuse esitada lühemalt:

$$P\{\hat{h}^2 < 0\} = P\{F_{a(n-1), a-1} > 1 + n\tau\}.$$

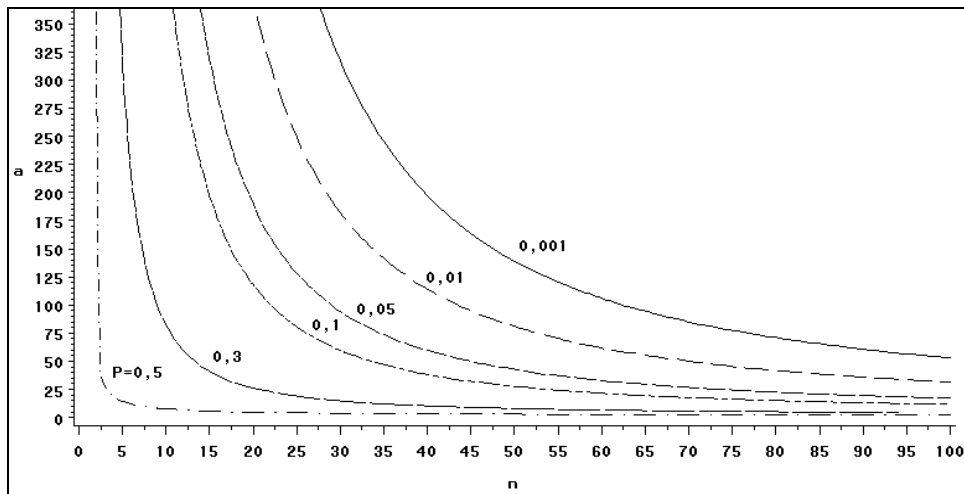
Analoogselt on tuletatav tõenäosus, et päritavuskoeffitsiendi hinnang on suurem ühest:

$$\begin{aligned} P\{\hat{h}^2 > 1\} &= P\left\{\frac{\hat{\sigma}_\sigma^2}{\hat{\sigma}_e^2} > \frac{1}{3}\right\} + \underbrace{P\{\hat{\sigma}_\sigma^2 < -\hat{\sigma}_e^2\}}_0 = P\left\{\frac{MSA}{MSE} > \frac{n}{3} + 1\right\} = \\ &= P\left\{F_{a(n-1), a-1} < \frac{3(1+n\tau)}{n+3}\right\}. \end{aligned}$$



Tabelis 3.3 on mõlemad tõenäosused arvatud parameetri  $\tau$  väärtustel 0,01; 0,05; 0,1 ja 0,25, isade arvu  $a$  väärtustel 2, 5, 10, 25 ja 50 ning tütarde arvu  $n$  väärtustel 5, 25 ja 50. Näiteks hinnates veiste piimatoodangu päritavust, mis Falconeri (1981) andmetel on ligikaudu 0,35, on tõenäosus illegaalse päritavuskoefitsiendi hinnangu saamiseks  $a = 2$  isa 50 järglase andmetel ( $n = 25$  tütar kummalgi isal) 0,54 (54%), 10 isa ja 10×25 tütre korral on vastav tõenäosus 0,09 (9%).

On selge, et nii isade kui ka tütarde arvu suurenemisel tõenäosus  $P_I$  väheneb. Seda illustreerivad  $(a, n)$ -tasandil kujutatud  $P_I = 0,5; 0,3; 0,1; 0,05; 0,01$  ja  $0,001$  graafikud erinevate  $\tau$  (ja  $h^2$ ) väärtuste korral (joonised 3.7, 3.8, 3.9 ja 3.10).



**Joonis 3.4.**  $P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\} = 0,5; 0,3; 0,1; 0,05; 0,01$  ja  $0,001$  graafikud  $\tau = 0,01$

( $\rho = 0,0099$  ja  $h^2 = 0,04$ ) korral;  $a$  - isade arv,  $n$  - tütarde arv isal.

**Tabel 3.3.** Tõenäosused  $P_I = P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\}$ , et päritavuskoeffitsiendi hinnang jääb teoreetilistest piiridest välja. Arvutused on tehtud tasakaalulisel juhul ( $a$  isa, igal isal  $n$  tüdart), eeldades, et uuritav tunnus on normaaljaotusega. Suurused  $\rho$  ja  $h^2$  avalduvad  $\tau$  abil valemitega (3.16) ja (3.18).

$\tau = 0,01$ ( $\rho = 0,0099$ ; $h^2 = 0,0396$ )									
$a$	$n = 5$			$n = 25$			$n = 100$		
	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$
2	0,6423	0,1497	0,7920	0,6244	0,0088	0,6332	0,5197	0,0001	0,5197
5	0,5453	0,0719	0,6171	0,4725	0,0000	0,4725	0,2642	0,0000	0,2642
10	0,5076	0,0208	0,5284	0,3834	0,0000	0,3834	0,1249	0,0000	0,1249
25	0,4664	0,0007	0,4671	0,2611	0,0000	0,2611	0,0204	0,0000	0,0204
50	0,4328	0,0000	0,4328	0,1631	0,0000	0,1631	0,0014	0,0000	0,0014

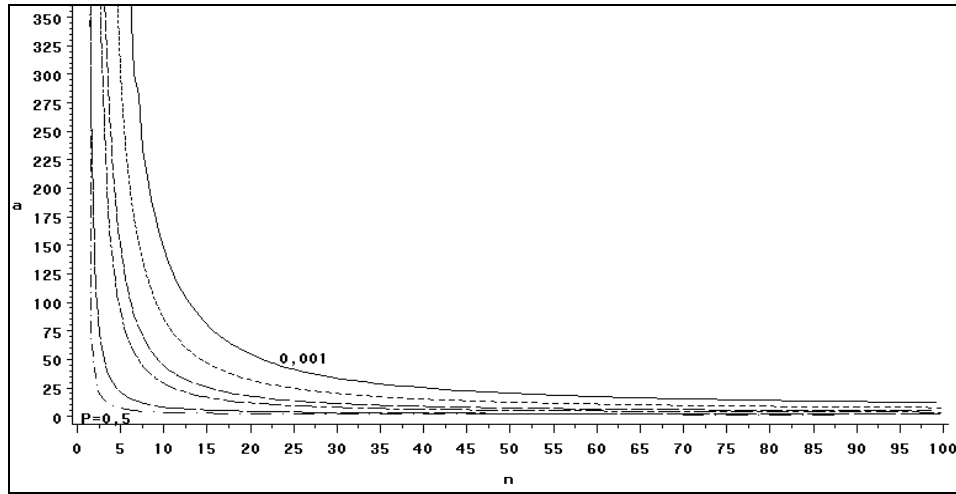
$\tau = 0,05$ ( $\rho = 0,0476$ ; $h^2 = 0,1905$ )									
$a$	$n = 5$			$n = 25$			$n = 100$		
	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$
2	0,6028	0,1823	0,7850	0,4918	0,0472	0,5390	0,3165	0,0177	0,3342
5	0,4606	0,1142	0,5749	0,2237	0,0035	0,2272	0,0447	0,0002	0,0449
10	0,3816	0,0490	0,4306	0,0902	0,0001	0,0903	0,0029	0,0000	0,0029
25	0,2713	0,0049	0,2761	0,0094	0,0000	0,0094	0,0000	0,0000	0,0000
50	0,1789	0,0001	0,1790	0,0003	0,0000	0,0003	0,0000	0,0000	0,0000

$\tau = 0,10$ ( $\rho = 0,0909$ ; $h^2 = 0,3636$ )									
$a$	$n = 5$			$n = 25$			$n = 100$		
	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$
2	0,5621	0,2191	0,7813	0,4046	0,1090	0,5136	0,2367	0,0788	0,3155
5	0,3774	0,1729	0,5504	0,1132	0,0356	0,1489	0,0147	0,0149	0,0296
10	0,2665	0,1031	0,3696	0,0218	0,0058	0,0275	0,0002	0,0010	0,0013
25	0,1278	0,0257	0,1535	0,0003	0,0000	0,0003	0,0000	0,0000	0,0000
50	0,0470	0,0031	0,0501	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

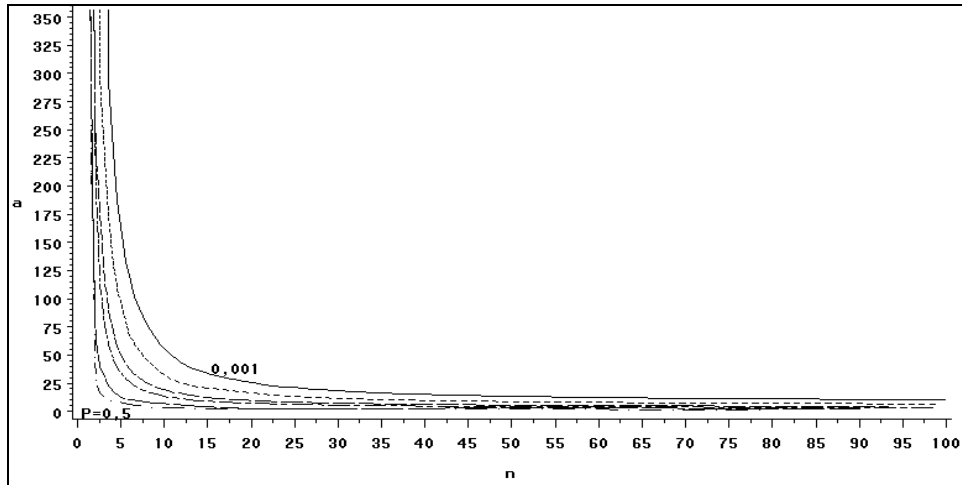
  

$\tau = 0,25$ ( $\rho = 0,2$ ; $h^2 = 0,8$ )									
$a$	$n = 5$			$n = 25$			$n = 100$		
	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$	$P\{\hat{h}^2 < 0\}$	$P\{\hat{h}^2 > 1\}$	$P_I$
2	0,4762	0,3080	0,7843	0,2880	0,2622	0,5502	0,1553	0,2519	0,4072
5	0,2249	0,3475	0,5724	0,0321	0,2788	0,3109	0,0028	0,2612	0,2640
10	0,0978	0,3307	0,4285	0,0014	0,2443	0,2457	0,0000	0,2216	0,2216
25	0,0126	0,2740	0,2866	0,0000	0,1634	0,1634	0,0000	0,1364	0,1364
50	0,0006	0,2089	0,2095	0,0000	0,0907	0,0907	0,0000	0,0669	0,0669

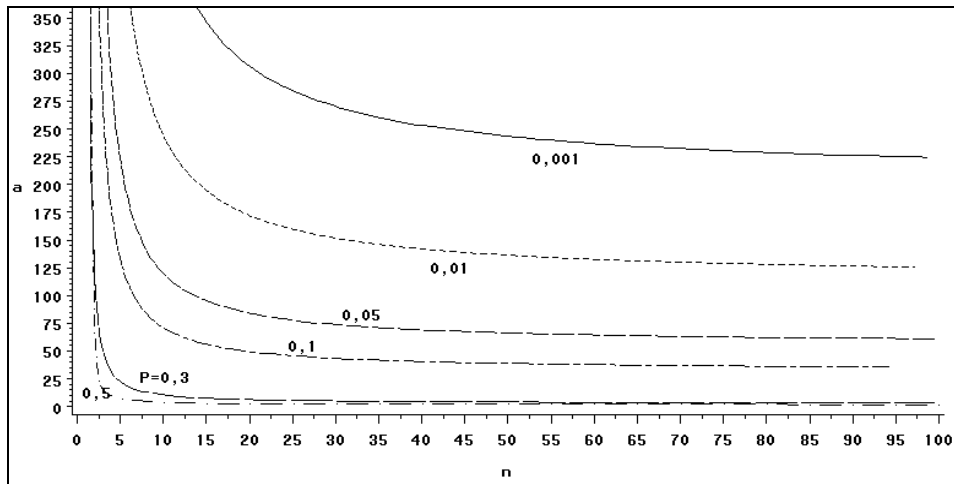


**Joonis 3.5.**  $P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\} = 0,5; 0,3; 0,1; 0,05; 0,01$  ja  $0,001$  graafikud  $\tau = 0,05$  ( $\rho = 0,0476$  ja  $h^2 = 0,19$ ) korral;  $a$  - isade arv,  $n$  - tütarde arv isal.

Nii väikeste kui ka suurte  $h^2$  väärtuste korral on jooned  $P_I$  graafikutel küllalt eraldatud (joonised 3.4 ja 3.7). Juhul  $\tau = 0,01$  on tegu eelkõige negatiivse päritavuskoeffitsiendi hinnangu saamise tõenäosusega,  $\tau = 0,25$  korral väljendab  $P_I$  aga ühest suurema päritavuskoeffitsiendi hinnangu saamise tõenäosust. Väga väikese päritavuse korral kahaneb hinnangu illegaalsuse tõenäosus kiiremini tütarde arvu kasvades isa kohta, suure päritavuse korral aga isade arvu kasvades. Vahepealsete päritavuste ( $h^2 = 0,19$  ja  $h^2 = 0,36$ ) korral on tõenäosuse  $P_I$  erinevatele väärtustele vastavad jooned  $(a, n)$ -tasandil üsna tihedalt koos ning seega väheneb illegaalse päritavuse hinnangu saamise tõenäosus kiiresti nii isade arvu kui ka tütarde arvu suurenemisel (joonised 3.5 ja 3.6). Tabelist 3.3 hakkab silma, et väga väikese isade ja tütarde arvu korral on tõenäosus saada negatiivset päritavuse hinnangut suurem kui ühest suurema hinnangu saamise tõenäosus, ja seda isegi päritavuse  $h^2 = 0,8$  juures (vt. ka Kaart, 1997).



**Joonis 3.6.**  $P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\} = 0,5; 0,3; 0,1; 0,05; 0,01$  ja  $0,001$  graafikud  $\tau = 0,1$  ( $\rho = 0,091$  ja  $h^2 = 0,36$ ) korral;  $a$  - isade arv,  $n$  - tütarde arv isal.



**Joonis 3.7.**  $P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\} = 0,5; 0,3; 0,1; 0,05; 0,01$  ja  $0,001$  graafikud  $\tau = 0,25$  ( $\rho = 0,2$  ja  $h^2 = 0,8$ ) korral;  $a$  - isade arv,  $n$  - tütarde arv isal.

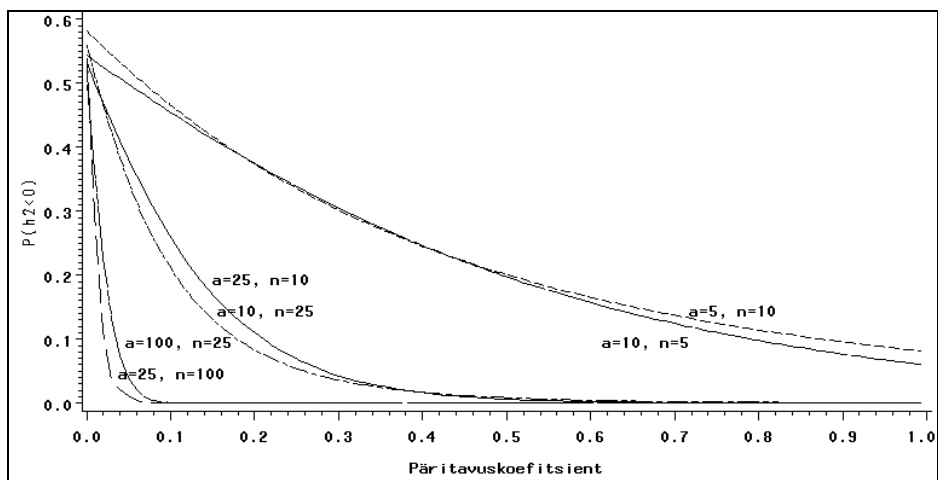
Saadud teoreetiliste tulemuste õigsust on kontrollitud statistilise modelleerimisega. Selleks genereeriti juhuslikud tasakaalulised andmestikud lähtudes tabelis 3.3 toodud isade ja tütarde arvudest ning päritavuskoeffitsiendi teoreetilistest väärtustest. Genereeritud andmetega teostati protseduuri ANOVA abil dispersioonanalüüs ja arvutati valemi (3.4) põhjal päritavuskoeffitsiendi väärtus. Kogu protsessi korrati SAS-makro abil 1000 korda, tulemused ühendati üheks failiks ja leiti sealt negatiivse päritavuskoeffitsiendi ja ühest suurema päritavuskoeffitsiendi hinnangute

suhtelised sagedused. Küllaltki suure ajakulu tõttu viidi modelleerimised läbi vaid mõningate variantide korral. Päritavuskoeffitsiendi illegaalsete hinnangute suhtelised sagedused (tabel 3.4) on küllalt lähedased tabelis 3.3 toodud teoreetilistele tõenäosustele.

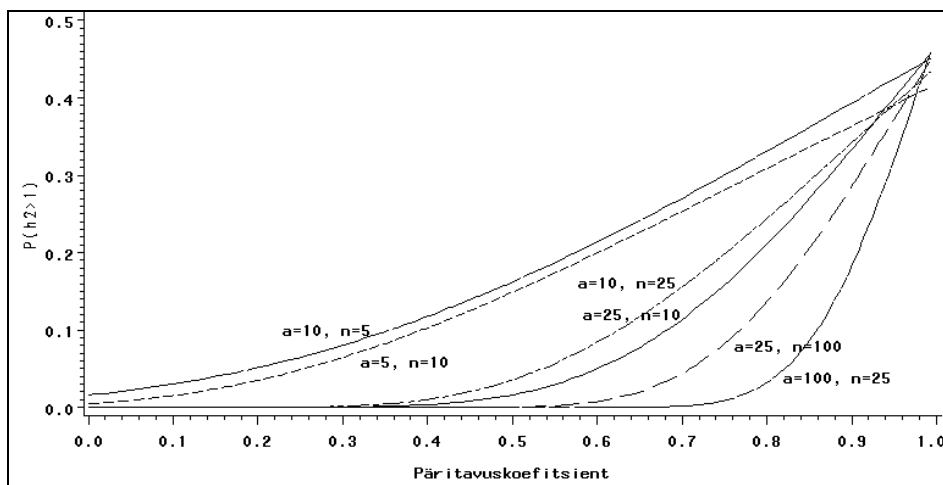
**Tabel 3.4.** Teoreetilised ja empiirilised tõenäosused  $P\{\hat{h}^2 < 0\}$ ,  $P\{\hat{h}^2 > 1\}$  ning  $P_I = P\{\hat{h}^2 < 0\} + P\{\hat{h}^2 > 1\}$ , et päritavuskoeffitsiendi hinnang jääb teoreetilistest piiridest välja. Empiirilised tõenäosused on leitud 1000 modelleerimise tulemuste alusel.

Isade arv (a) ja tütarde arv isal (n)	Päritavus- koeffitsiendi väärtus	$P\{\hat{h}^2 < 0\}$		$P\{\hat{h}^2 > 1\}$		$P_I$	
		Teoreetiline e	Empiiriline ne	Teoreetiline ne	Empiiriline ne	Teoreetiline ne	Empiiriline ne
$a = 5, n = 5$	0,0396	0,5453	0,5990	0,0719	0,0640	0,6171	0,6630
	0,1905	0,4606	0,4840	0,1142	0,1300	0,5749	0,6140
	0,3636	0,3774	0,3700	0,1729	0,1700	0,5504	0,5400
	0,8	0,2249	0,2450	0,3475	0,3210	0,5724	0,5660
$a = 10, n = 25$	0,0396	0,3834	0,3820	0,0000	0,0000	0,3834	0,3820
	0,1905	0,0902	0,1010	0,0001	0,0000	0,0903	0,1010
	0,3636	0,0218	0,0250	0,0058	0,0040	0,0275	0,0290
	0,8	0,0014	0,0010	0,2443	0,2310	0,2457	0,2320
$a = 25, n = 100$	0,0396	0,0204	0,0130	0,0000	0,0000	0,0204	0,0130
	0,1905	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	0,3636	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	0,8	0,0000	0,0000	0,1364	0,0220	0,1364	0,0220

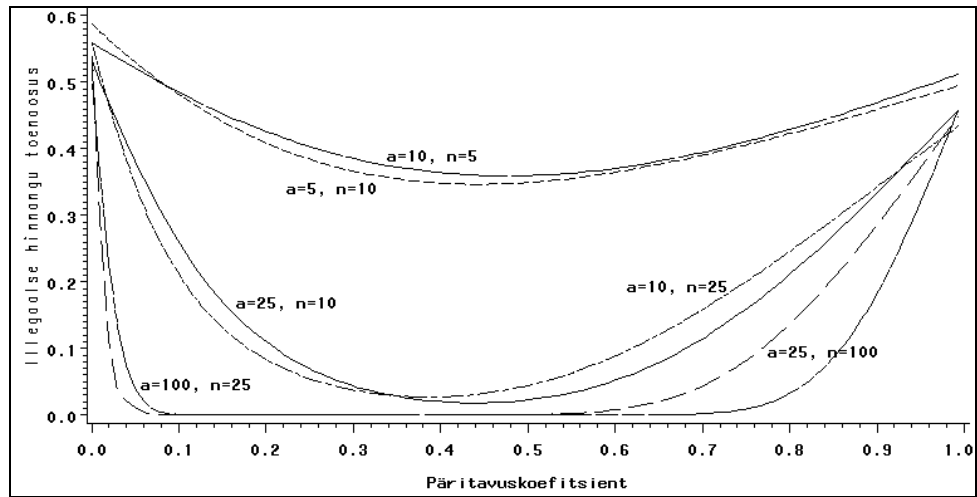
Päritavuskoeffitsiendi illegaalse hinnangu tõenäosuse seost päritavuskoeffitsiendi teoreetilise väärtusega ning isade ja tütarde arvu vahekorraga illustreerivad joonised 3.8, 3.9 ja 3.10. Negatiivse  $\hat{h}^2$  tõenäosus kasvab päritavuskoeffitsiendi väärtuse vähenedes kiiremini (järsemalt) kui ühest suurema  $\hat{h}^2$  tõenäosus päritavuskoeffitsiendi väärtuse kasvades. Tütarde arvu suurenemine isa kohta vähendab negatiivse päritavuskoeffitsiendi hinnangu tõenäosust kiiremini, võrreldes isade arvu suurenemisega. Seevastu ühest suurema päritavuskoeffitsiendi hinnangu tõenäosus kahaneb jõudsamalt isade arvu kasvades.



**Joonis 3.8.** Päritavuskoeffitsiendi negatiivse hinnangu tõenäosus erinevate  $h^2$  väärtuste ning isade arvu  $a$  ja tütarde arvu  $n$  korral.



**Joonis 3.9.** Päritavuskoeffitsiendi ühest suurema hinnangu tõenäosus erinevate  $h^2$  väärtuste ning isade arvu  $a$  ja tütarde arvu  $n$  korral.



**Joonis 3.10.** Päritavuskoeffitsiendi illegaalse hinnangu tõenäosus erinevate  $h^2$  väärtuste ning isade arvu  $a$  ja tütarde arvu  $n$  korral.

## 4. Dispersioonikomponentide hindamine mittetasakaalulisel juhul

Mittetasakaalulisel juhul on dispersioonikomponentide ja seeläbi ka päritavuse hindamiseks välja töötatud hulk meetodeid, mis arvutuslikult on küllalt keerukad ja võivad anda erinevaid resultate samade mudelite ja andmete korral. Parimat meetodit ei leidu ja seetõttu leiavad üsna paljud neist laialdast rakendust.

Järgnevalt kirjeldatavate meetodite kompaksemaks esitamiseks tähistame

$$W = [X \quad Z] \quad (4.1)$$

ja

$$\beta = [b' \quad u']'. \quad (4.2)$$

Normaalvõrrandid (2.6) on siis kirjutatavad kujul

$$W'W\beta = W'y. \quad (4.3)$$

### 4.1. Hendersoni meetodid

Hendersoni kolm meetodit, avaldatuna 1953. aastal (Henderson, 1984), olid esimesed meetodid dispersioonikomponentide leidmiseks mittetasakaaluliste andmete korral. Kõik nad on tasakaalulisel juhul kasutatava ANOVA kohandused, annavad nihketa hinnangud, kuid ei välista negatiivseid hinnanguid. Tasakaalulisel juhul kasutatuna on nende abil saadud hinnangud võrdsed ANOVA abil saadud hinnangutega. Kuna ükski neist meetoditest ei ole iteratiivne, ei vaja nad mingit eelinformatsiooni dispersioonikomponentide hindamiseks ja arvutusprotsess kulgeb suhteliselt kiirelt.

Hendersoni 1. meetod on ilmselt lihtsaim meetod hindamiseks dispersioonikomponente mittetasakaaluliste andmete korral. Ta on kasutatav vaid täiesti juhusliku mudeli korral, s.t.  $Xb = 1\mu$ . Meetod baseerub ruutvormidel, mis on analoogsed tasakaalulisel juhul kasutatutega. Tänu andmete mittetasakaalulisusele ei pruugi need ruutvormid olla alati mittenegatiivselt määratud ja seega ei saa neid esitada ruutude summana (Searle jt., 1992; Henderson, 1984).

Hendersoni 2. meetod on konstrueeritud lähtudes soovist kasutada 1. meetodi lihtsat algoritmi segamudelite korral. Meetod koosneb kahest osast. Esiteks loetakse kõik juhuslikud efektid mudelist (2.1) fikseerituteks ja lahendatakse normaalvõrrandid (2.6) parameetri  $b$  suhtes. Seejärel hinnatakse dispersioonikomponente 1. meetodiga mudelist:



$$y^* = y - X\tilde{b}.$$

Meetod on kasutatav segamudelite korral, kus puuduvad interaktsioonid fikseeritud ja juhuslike faktorite vahel (Searle jt., 1992; Henderson, 1984).

Hendersoni 3. meetod baseerub üldisel lineaarsel mudelil (2.1) ja selle sobivalt valitud alammudelitel normaalvõrrandite (2.6) abil konstrueeritud ruutvormidel. Tänu vastavatele arvutiprogrammidele (Harvey, 1990) leidis meetod juba 60-ndatel aastatel ja leiab ka tänapäeval laialdast kasutust.

Mudelist, mis sisaldab  $s$  erinevat juhuslikku faktorit, on võimalik leida  $2^s$  alammudelit, millest tundmatute dispersioonikomponentide hindamiseks on vajalikud vaid  $s+2$ . Seega ei ole Hendersoni 3. meetodiga saadavad hinnangud ühesed, vaid sõltuvad alammudelite valikust. On loomulik teha valik, mis garanteerib parameetrite hinnangute vähima varieeruvuse, kuid selline alammudelite kombinatsioon on teadmata. Seetõttu kasutatakse meetodis kokkuleppeliselt lihtsaimaid variante. Kasutatavad alammudelid ja ruutvormid on:

1) konstantne mudel, vastavaks ruutvormiks on  $y'y$ , s.o. ruutude summa;

2) täielik mudel, ruutvorm on kujul  $\tilde{\beta}'W'y$ , kus  $\tilde{\beta}$  on parameetervektori (4.2) vähimruutude hinnang normaalvõrrandist (4.3);

3) mudel, mis sisaldab vaid juhuslikku efekti  $u_i$ , mille dispersiooni  $\sigma_i^2$  soovitakse hinnata, ruutvorm on esitatav kujul  $\tilde{u}_i'Z_i'My$ , kus  $\tilde{u}_i$  on juhusliku efekti  $u_i$  vähimruutude hinnang ja

$$M = I - X(X'X)^{-1}X' \quad (4.4)$$

on nn. taseme tsentreerimismatriks (*level-centering matrix*), mis absorbeerib fikseeritud efektide mõju;  $MX = 0$ , maatriks  $M$  on idempotentne ja sümmeetriline (Möls, Parring, 1994).

Juhusliku vea hinnang on tähistuste (4.1) ja (4.2) abil leitav mudelist (2.1) seosega  $\tilde{e}' = y' - \tilde{\beta}'W'$ . Dispersioonikomponendi  $\sigma_e^2$  hindamiseks kasutatav ruutvorm avaldub esimesest ja teisest alammudelist:

$$\tilde{e}'\tilde{e} = y'y - \tilde{\beta}'W'y, \quad (4.5)$$

sest  $y'W\tilde{\beta} - \tilde{\beta}'W'W\tilde{\beta} = 0$ .

Ruutvormi (4.5) keskvärtus kohal  $i = 0$  valemis (2.5) on punktis 2.2 toodud ruutvormi keskvärtuse valemi alusel

$$E(y'y - \tilde{\beta}'W'y) = b'X'Xb + N\sigma_e^2 - b'X'W(W'W)^-W'Xb - tr(W(W'W)^-W')\sigma_e^2. \quad (4.6)$$

Viimane võrdus lihtsustub, arvestades maatriksi  $W$  kuju (4.1) ning üldistatud pöördmaatriksite teooriast teadaolevaid tulemusi:

$$b'X'W(W'W)^-W'Xb = b'X'Xb$$

ja

$$tr(W(W'W)^-W') = tr(WW^+),$$

kus  $W^+ = W'(WW')^-W(W'W)^-W'$  on Moore-Penrose üldistatud pöördmaatriks.

Korrutis  $WW^+$  on idempotentne ja seega võrdub tema jälg astakuga:

$$tr(WW^+) = r(WW^+) = r(W) = r(X) + r(Z) - 1$$

(Searle, 1987). Võrdus (4.6) saab kuju

$$E(y'y - \tilde{\beta}'W'y) = N\sigma_e^2 - r(W)\sigma_e^2,$$

kust juhuslikule veale vastav dispersioonikomponent on hinnatav valemiga

$$\hat{\sigma}_e^2 = \frac{y'y - \tilde{u}'Z'y - \tilde{b}'X'y}{N - r(X) - r(Z) + 1}. \quad (4.7)$$

Ülejäänud dispersioonikomponentide hindamiseks esitame normaalvõrrandid (2.6) maatriksi  $M$  abil:

$$\begin{bmatrix} Z_1'MZ_1 & \cdots & Z_1'MZ_s \\ \vdots & \ddots & \vdots \\ Z_s'MZ_1 & \cdots & Z_s'MZ_s \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_s \end{bmatrix} = \begin{bmatrix} Z_1'My \\ \vdots \\ Z_s'My \end{bmatrix}.$$

Siit  $\tilde{u}_i = (Z_i'MZ_i)^-Z_i'My$  ning

$$\tilde{u}_i'Z_i'My = y'M'Z_i(Z_i'MZ_i)^-Z_i'My. \quad (4.8)$$

Tähistades  $\mathcal{E}_i = (Z_i'MZ_i)^-$  ning kasutades dispersioonimaatriksi  $V$  esitust (2.4), on ruutvormi (4.8) keskvärtus punktis 2.2 toodud ruutvormi keskvärtuse valemi alusel

$$E(\tilde{u}_i'Z_i'My) = \sum_{j=1}^s tr[\mathcal{E}_i(Z_i'MZ_j)(Z_j'MZ_i)]\sigma_j^2 + tr(\mathcal{E}_iZ_i'MZ_i)\sigma_e^2. \quad (4.9)$$

Üldistatud pöördmaatriksite teooriast on teada, et  $(Z'MZ)(Z'MZ)^-(Z'MZ) = (Z'MZ)$ , mistõttu

$$tr[\mathcal{E}_i(Z_i'MZ_j)(Z_j'MZ_i)] = tr(Z_i'MZ_i), \quad i = j \text{ korral.}$$

Arvestades, et  $(Z'MZ)^-(Z'MZ)$  on idempotentne maatriks astakuga

$$r(Z'MZ) = r(MZ) = r\begin{bmatrix} X \\ Z \end{bmatrix} - r(X) = r(Z) - 1,$$

ning teades, et idempotentse maatriksi jälg on võrdne tema astakuga, saab  $\sigma_e^2$  kordaja valemis (4.9) kujul

$$tr(\mathcal{L}_i Z_i' M Z_i) = r(Z_i) - 1.$$

Ühe juhusliku faktoriga (isa) mudeli korral on vastav dispersioonikomponent hinnatav valemist

$$\hat{\sigma}_e^2 = \frac{\tilde{u}' Z' M y - (r(Z) - 1) \sigma_e^2}{tr(Z'MZ)}. \quad (4.10)$$

## 4.2. MIVQUE meetod

Tasakaalulisel juhul on ANOVA hinnangud dispersioonikomponentidele minimaalse varieeruvusega invariantseid nihketa hinnangud. Tasakaalustamata juhul on minimaalse varieeruvusega hinnangud leitavad vaid teadaolevate dispersioonimaatriksite  $G$  ja  $R$  korral. Meetod, mis võimaldab leida lokaalselt parimad invariantseid nihketa hinnangud, kannab normaalsuse eeldustel nime MIVQUE (*minimum variance quadratic unbiased estimation*). „Lokaalselt parim” tähendab, et MIVQUE hinnangud on vähima varieeruvusega hinnangud invariantsete nihketa hinnangute klassis konkreetsete etteantud dispersioonimaatriksite korral. Meetodi töötasid 1970. aastaks teineteisest sõltumatult välja C. R. Rao ja L. R. LaMotte. Segamudeli võrrandil (2.7) baseeruvad arvutusvalemid tuletas 1973. aastal C. R. Henderson. Siinse tuletuskäigu „selgroog” pärineb Mao ja Schaeffer (1993) loengukonspektist.

Eeldame, et  $y \sim N(Xb, V)$  ning tähistame  $\sigma_e^2 = \sigma_0^2$ . Olgu  $\theta_i$  meie alglähend  $\sigma_i^2$  kohta,  $i = 0, \dots, s$ , ja  $\dot{V}$  meie alglähend  $V$  kohta. Siis, lähtudes seostest (2.4) ja (2.5), avaldub  $\dot{V}$  kujul

$$\dot{V} = \sum_{i=1}^s Z_i Z_i' \theta_i + I \theta_0 = \sum_{i=0}^s V_i \theta_i.$$

LaMotte tõestas, et ruutvormi

$$(y - X\tilde{b})' \dot{V}^{-1} Z_i Z_i' \dot{V}^{-1} (y - X\tilde{b}) \quad (4.11)$$

võrdsustamine oma ooteväärtusega annab MIVQUE hinnangu dispersiooni-  
komponendile  $\sigma_i^2$ ,  $\tilde{b} = (X\dot{V}^{-1}X)^- X\dot{V}^{-1}y$  on siin fikseeritud efektide üldine  
vähimruutude hinnang kohal  $V = \dot{V}$  (Henderson, 1984).

Tähistame

$$P = \dot{V}^{-1} - \dot{V}^{-1}X(X\dot{V}^{-1}X)^- X\dot{V}^{-1}, \quad (4.12)$$

kus  $P$  on projektsioonimaatriks,  $PX = 0$  ja  $P\dot{V}$  on idempotentne (Mao, Schaeffer,  
1993). Siis saab korrutise  $Py$  avaldada kujul

$$\begin{aligned} Py &= [\dot{V}^{-1} - \dot{V}^{-1}X(X\dot{V}^{-1}X)^- X\dot{V}^{-1}]y = \\ &= \dot{V}^{-1}[y - X(X\dot{V}^{-1}X)^- X\dot{V}^{-1}y] = \dot{V}^{-1}(y - X\tilde{b}), \end{aligned}$$

millest

$$(y - X\tilde{b})'\dot{V}^{-1}Z_i Z_i \dot{V}^{-1}(y - X\tilde{b}) = y'PZ_i Z_i'Py. \quad (4.13)$$

Arvestades, et  $\hat{u}_i = \dot{G}_i Z_i \dot{V}^{-1}(y - X\tilde{b})$ , kus  $\hat{u}_i$  on juhusliku faktori  $u_i$  parim lineaarne  
nihketa prognoos (2.10) ja  $\dot{G}_i = I_{n_i} \theta_i$  on  $u_i$  eeldatav dispersioonimaatriks, saame jätkata

$$y'PZ_i Z_i'Py = (y - X\tilde{b})'\dot{V}^{-1}Z_i \dot{G}_i (\dot{G}_i^{-1} \dot{G}_i^{-1}) \dot{G}_i Z_i \dot{V}^{-1}(y - X\tilde{b}) = \hat{u}_i' \dot{G}_i^{-2} \hat{u}_i = \hat{u}_i' \hat{u}_i / \theta_i^2. \quad (4.14)$$

Jääkvarieeruvuse hindamiseks kasutatav ruutvorm saadakse seostest (4.12 - 4.14),  
võttes  $i = 0$ . Siis

$$y'PPy = \hat{e}'\hat{e} / \theta_0^2 = (y'y - \tilde{b}'X'y - \sum_{i=1}^s \hat{u}_i' Z_i' y - \sum_{i=1}^s \hat{u}_i' \hat{u}_i k_i) / \theta_0^2, \quad (4.15)$$

kus  $k_i = \theta_0 / \theta_i$  (Henderson, 1984).

Seega on MIVQUE hinnangu saamiseks vajalik võrdsustada oma ooteväärtustega  
ruutvormid

$$\hat{u}_i' \hat{u}_i \text{ ja } y'y - \tilde{b}'X'y - \sum_{i=1}^s \hat{u}_i' Z_i' y.$$

Ruutvormide ooteväärtuste leidmiseks kirjutame esmalt Hendersoni segamudeli  
võrrandi (2.8) välja pikemalt:

$$\begin{bmatrix} X'X & X'Z_1 & \dots & X'Z_s \\ Z_1'X & Z_1'Z_1 + Ik_1 & \dots & Z_1'Z_s \\ \vdots & \vdots & \ddots & \vdots \\ Z_s'X & Z_s'Z_1 & \dots & Z_s'Z_s + Ik_s \end{bmatrix} \begin{bmatrix} \tilde{b} \\ \hat{u}_1 \\ \vdots \\ \hat{u}_s \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ \vdots \\ Z_s'y \end{bmatrix}.$$

Tähistame selle võrrandi kordajate (ruutmaatriksi) üldistatud pöördmaatriksi

$$C = \begin{bmatrix} C_{xx} & C_{x1} & \dots & C_{xs} \\ C_{1x} & C_{11} & \dots & C_{1s} \\ \vdots & \vdots & \ddots & \vdots \\ C_{sx} & C_{s1} & \dots & C_{ss} \end{bmatrix} = \begin{bmatrix} C_x \\ C_1 \\ \vdots \\ C_s \end{bmatrix} \quad (4.16)$$

ning segamudeli võrrandi parema poole

$$r = W'y = [X \quad Z_1 \quad \dots \quad Z_s]' y.$$

Siis on vektori  $r$  dispersioon, lähtudes punktis 2.2 toodud ruutvormide dispersiooni valemist, avaldatav kujul

$$D(r) = W'D(y)W = \sum_{j=0}^s W'V_jW\sigma_j^2 = \sum_{j=1}^s (W'Z_j)(Z_j'W)\sigma_j^2 + W'W\sigma_0^2 = \sum_{j=0}^s B_j\sigma_j^2,$$

kus  $B_j = W'Z_jZ_j'W$  ning  $B_0 = W'W$ . Kasutades sissetoodud tähistusi, on juhusliku efekti  $u_i$  prognoos avaldatav seosest

$$\hat{u}_i = C_i r,$$

millest

$$\hat{u}_i' \hat{u}_i = r' C_i' C_i r.$$

Seega on ruutvormi  $\hat{u}_i' \hat{u}_i$  ooteväärtus punkti 2.2 alusel

$$E(\hat{u}_i' \hat{u}_i) = \sum_{j=0}^s \text{tr}(C_i' C_i B_j) \sigma_j^2.$$

Sarnaselt saame, et

$$E(y'y) = b'X'Xb + N \sum_{j=0}^s \sigma_j^2$$

ja

$$\tilde{b}'X'y + \sum_{i=1}^s \hat{u}_i' Z_i' y = r'Cr.$$

Siis

$$E(r'Cr) = b'X'WCW'Xb + \sum_{j=0}^s \text{tr}(CB_j) \sigma_j^2,$$

kusjuures  $X'WCW'X = X'X$  (Mao, Schaeffer, 1993). Nüüd on teise vajamineva ruutvormi ooteväärtus väljendatav valemiga

$$E(y'y - r'Cr) = \sum_{j=0}^s (N - \text{tr}(CB_j)) \sigma_j^2.$$

Ühe juhusliku faktoriga (isa) mudeli korral hinnatakse dispersioonikomponente maatriksvõrdusest

$$\begin{bmatrix} \hat{\sigma}_\sigma^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} \text{tr}(C'_\sigma C_\sigma W' Z Z' W) & \text{tr}(C'_\sigma C_\sigma W' W) \\ N - \text{tr}(C W' Z Z' W) & N - \text{tr}(C W' W) \end{bmatrix}^{-1} \begin{bmatrix} \hat{u}' \hat{u} \\ y' y - y' W C W' y \end{bmatrix}, \quad (4.17)$$

kus  $C_\sigma$  tähistab juhuslikule faktorile isa vastavaid ridu üldistatud pöördmaatriksis  $C$  (4.16).

### 4.3. MIVQUE(0) meetod

Enam kasutust leiab MIVQUE(0) meetod, s.o. lihtsustatud MIVQUE meetod, kus alglähendina dispersioonikomponentidele, v.a. juhuslikule veale vastav dispersioonikomponent, kasutatakse nulle. Meetod leidub mitmetes statistikapakettides, nagu näiteks SAS, Harvey programm (Harvey, 1990).

Olgu  $\theta_i = 0$  meie alglähend  $\sigma_i^2$  kohta,  $i = 1, \dots, s$ , ja  $\theta_0$  meie alglähend  $\sigma_e^2$  kohta. Siis esitub dispersioonimaatriks  $\dot{V}$  kujul

$$\dot{V} = I\theta_0$$

ning fikseeritud efektid on hinnatavad valemist

$$\tilde{b} = (X'X)^{-1} X'y.$$

MIVQUE meetodis kasutatud projektsioonimaatriks  $P$  (4.12) võrdub seosega (4.4) defineeritud projektsioonimaatriksiga  $M$ ,

$$My = y - X\tilde{b}$$

ja võrdus (4.13) on üleskirjutatav seosena

$$(y - X\tilde{b})' \dot{V}^{-1} Z_i Z_i' \dot{V}^{-1} (y - X\tilde{b}) = y' M' Z_i Z_i' M y / \theta_0^2.$$

Juhuslikule veale vastava dispersioonikomponendi  $\sigma_e^2$  hindamiseks kasutatav ruutvorm on nüüd kujul

$$y'y - \tilde{b}' X'y.$$

Ruutvormide ooteväärtused on:

$$E(y' M' Z_i Z_i' M y) = \sum_{j=1}^s \text{tr} \left[ (Z_i' M Z_j) (Z_j' M Z_i) \right] \sigma_j^2 + \text{tr}(Z_i' M Z_i) \sigma_e^2$$

ja

$$E(y'y - \tilde{b}' X'y) = \sum_{j=1}^s \left[ N - \text{tr}(Z_j' X (X'X)^{-1} X' Z_j) \right] \sigma_j^2 + \text{tr}(M) \sigma_e^2.$$

Ühe juhusliku faktoriga (isa) mudeli korral avalduvad dispersioonikomponendid järgmise maatriksvõrdusega:

$$\begin{bmatrix} \hat{\sigma}_\sigma^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} \text{tr}[(Z'MZ)(Z'MZ)] & \text{tr}(Z'MZ) \\ \text{tr}(Z'MZ) & \text{tr}(M) \end{bmatrix}^{-1} \begin{bmatrix} y'M'ZZ'My \\ y'y - \tilde{b}'X'y \end{bmatrix}. \quad (4.18)$$

#### 4.4. ML meetod

Suurima tõepära (ML) meetodi kasutamist dispersioonikomponentide hindamisel kirjeldasid 1967. aastal Hartley ja Rao. Sellest ajast alates on meetod olnud laialt kasutatav ja sisaldub enamikes dispersioonikomponente mittetasakaalulisel juhul hinnata võimaldavates arvutiprogrammides. Samuti baseeruvad selle meetodi või järgnevas peatükis kirjeldatava REML meetodi modifikatsioonidel mitmed spetsiaalsed aretuses kasutatavad programmid (Boldman jt., 1995; Groeneveld, 1995).

Eeldame, et andmed on kirjeldatud üldise lineaarse mudeliga (2.1), on peatükis 2 toodud dispersioonistruktuuriga (2.2 - 2.5) ja  $y \sim N(Xb, V)$ . Siis avaldub uuritava tunnuse  $y$  logaritmiline tõepärafunktsioon kujul

$$l(b, V; y) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |V| - \frac{1}{2} (y - Xb)'V^{-1}(y - Xb).$$

Logaritmilise tõepärafunktsiooni tuletised parameetrite  $b$  ja  $\sigma_i^2$ ,  $i = 0, \dots, s$ , järgi on

$$\frac{\partial l(b, V; y)}{\partial b} = -X'V^{-1}Xb + X'V^{-1}y$$

ja

$$\frac{\partial l(b, V; y)}{\partial \sigma_i^2} = \frac{1}{2} \text{tr}(V^{-1}V_i) + \frac{1}{2} (y - Xb)'V^{-1}V_iV^{-1}(y - Xb),$$

sest  $\frac{\partial V}{\partial \sigma_i^2} = Z_i Z_i' = V_i$ . Võrdsustades tuletised nulliga, saame

$$\tilde{b} = (X'V^{-1}X)^{-1} X'V^{-1}y \quad (4.19)$$

ja

$$\text{tr}(V^{-1}V_i) = (y - X\tilde{b})'V^{-1}V_iV^{-1}(y - X\tilde{b}). \quad (4.20)$$

Võttes kasutusele valemiga (4.12) defineeritud projektsioonimaatriksi  $P$  ( $V = \dot{V}$ ) ning arvestades seost  $Py = V^{-1}(y - X\tilde{b})$ , on võrdusele (4.20) antav alternatiivne kuju

$$\text{tr}(V^{-1}Z_i Z_i') = y'PV_iPy. \quad (4.21)$$

Võrdused (4.20) ja (4.21) pole üldjuhul lineaarsed parameetri  $\sigma_i^2$  suhtes. Nad lahenduvad iteratiivselt ning saadavad hinnangud on nihkega (pole täidetud punktis 2.2 toodud hinnangute nihutamatus tingimused).

Dispersioonikomponendi  $\sigma_e^2$  hindamiseks tuleb lahendada võrrand (4.21) kohal  $i = 0$ . Võrrandi parem pool  $y'PPy$  esitub seosena (4.15), vasak pool avaldub kujul

$$tr(P) = N\sigma_e^{-2} - \left(\sum_{i=1}^s \hat{u}_i' \hat{u}_i k_i\right) \sigma_e^{-4} \quad (4.22)$$

(Mao, Schaeffer, 1993). Võrdsustades seosed (4.15) ja (4.22), saame

$$\hat{\sigma}_e^2 = (y'y - \tilde{b}'X'y - \hat{u}'Z'y) / N. \quad (4.23)$$

Dispersioonikomponentide  $\sigma_i^2, i = 1, \dots, s$ , hindamiseks võtame kasutusele  $(n \times n)$ -maatriksid

$$T = (Z'R^{-1}Z + G^{-1})^{-1} \quad (4.24)$$

ja  $F_{ii}$ . Maatriks  $F_{ii}$  on dispersioonimaatriksi  $G$  struktuuriga, kus  $\sigma_i^2$  kohal on ühed ja  $\sigma_j^2, i \neq j$ , kohal nullid. Siis  $GF_{ii} = \sigma_i^2 F_{ii}$  ja  $F_{ii} = GF_{ii} / \sigma_i^2$ . Arvestades, et dispersioonimaatriksi  $V$  pöördmaatriks  $V^{-1}$  on avaldatav kujul

$$V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \quad (4.25)$$

(Harville, 1977), on võrduse (4.21) vasak pool pikemalt väljakirjutatav seosena

$$\begin{aligned} tr(V^{-1}V_i) &= tr\left[(R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1})Z_iZ_i'\right] = \\ &= tr\left[(R^{-1} - R^{-1}ZTZ'R^{-1})ZF_{ii}Z'\right]. \end{aligned} \quad (4.26)$$

Dispersioonimaatriksi  $R$  esituse (2.3) tõttu on maatriksi  $T$  antav kujul

$$T = (Z'Z + G^{-1}\sigma_e^2)^{-1}\sigma_e^2. \quad (4.27)$$

Siit

$$T\sigma_e^{-2}(Z'Z + G^{-1}\sigma_e^2) = I_n,$$

millest

$$TZ'Z\sigma_e^{-2} = I_n - TG^{-1}.$$

Kasutades viimast seost ja maatriksi  $R$  esitust (2.3), jätkame võrduste ahelat (4.26):

$$\begin{aligned} tr(V^{-1}Z_iZ_i') &= tr\left[(Z'Z - Z'ZTZ'Z\sigma_e^{-2})F_{ii}\sigma_e^{-2}\right] = \\ &= tr\left\{\left[Z'Z - Z'Z(I_n - TG^{-1})\right]F_{ii}\sigma_e^{-2}\right\} = tr(Z'ZTG^{-1}F_{ii}\sigma_e^{-2}) = \end{aligned}$$



$$= \text{tr}[(I_n - TG^{-1})G^{-1}F_{ii}].$$

Arvestades, et  $T$  on jagatav alammaatriksiteks vastavalt juhuslikele faktoritele:

$$T = \begin{bmatrix} T_{11} & \cdots & T_{1s} \\ \vdots & \ddots & \vdots \\ T_{s1} & \cdots & T_{ss} \end{bmatrix},$$

ning dispersioonimaatriks  $G$  on esitatav maatriksite otsesummana:

$$G = \bigoplus_{i=1}^s I_{n_i} \sigma_i^2,$$

avalduub võrduse (4.21) vasak pool lõpuks maatriksi  $F_{ii}$  kujust lähtudes seosena

$$\text{tr}(V^{-1}Z_i Z_i') = \text{tr}(I_{n_i} - T_{ii} \sigma_i^{-2}) \sigma_i^{-2} = n_i \sigma_i^{-2} - \text{tr} T_{ii} \sigma_i^{-4}. \quad (4.28)$$

Valemi (4.20) parem pool esitub analoogselt võrdusega (4.14), võttes  $V = \dot{V}$  ja  $G = \dot{G}$ :

$$y' P V_i P y = \hat{u}_i' \hat{u}_i / \sigma_i^4. \quad (4.29)$$

Võrdsustades seosed (4.28) ja (4.29), saame dispersioonikomponentide hinnangud

$$\hat{\sigma}_i^2 = (\hat{u}_i' \hat{u}_i + \text{tr} T_{ii}) / n_i, \text{ kus } i = 1, \dots, s.$$

#### 4.5. REML meetod

Suurima tõepära hinnangu puuduseks on see, et meetod ei võta arvesse fikseeritud efektide hindamiseks kasutatud vabadusastmete arvu. Näiteks  $n$  sõltumatu sama jaotusega  $N(\mu, \sigma^2)$  vaatluse  $x_1, \dots, x_n$  dispersiooni nihketa hinnang on  $\hat{\sigma}^2 = \sum_i (x_i - \bar{x})^2 / (n-1)$ , ML hinnang aga  $\hat{\sigma}_{ML}^2 = \sum_i (x_i - \bar{x})^2 / n$ . Suurima tõepära meetodi modifikatsioon, mis arvestab vabadusastmete kadu, kannab nime REML (*restricted* või *residual maximum likelihood method*). Patterson ja Thompson tuletasid meetodi 1971. aastal, kasutades nn. vea kontraste, milleks on vektori  $y$  lineaarkombinatsioonid saaduna mudeli (2.1) parandamisel fikseeritud efektide suhtes. Mudel (2.1) esitatakse kujul

$$K'y = K'Xb + K'Zu, \quad (4.30)$$

mis ei sisalda enam fikseeritud efekte –  $K'X = 0$ . Siis  $K'y \sim N(0, K'VK)$ , kusjuures  $r(K') = N - r(X)$ .

Võrrandid REML hinnangute saamiseks on tuletatavad ML võrrandeist (4.20), tehes asendused:  $y \rightarrow K'y$ ,  $Z \rightarrow K'Z$ ,  $X \rightarrow K'X = 0$  ja  $V \rightarrow K'VK$ . Siis on REML hinnangud leitavad võrrandeist

$$\text{tr}[(K'VK)^{-1}K'Z_iZ_i'K] = y'K(K'VK)^{-1}K'Z_iZ_i'K(K'VK)^{-1}K'y. \quad (4.31)$$

Searle (1992) tõestab, et kui  $r(K') = N - r(X)$ ,  $K'X = 0$  ja  $V$  on positiivselt määratud, siis

$$K(K'VK)^{-1}K' = P, \quad (4.32)$$

kus  $P$  on seosega (4.12) defineeritud projektsioonimaatriks ( $V = \dot{V}$ ). Arvestades nüüd valemist (4.30), on REML hinnangute võrrandid (4.31) esitatavad kujul

$$\text{tr}(PZ_iZ_i') = y'PZ_iZ_i'Py. \quad (4.33)$$

Analoogselt ML meetodile, pole ka REML hinnangud nihketa ning seos (4.33) on lahendatav iteratiivselt. REML meetodi puuduseks on asjaolu, et sõltuvalt mudeli (4.30) kujust, ei ole meetodiga võimalik hinnata fikseeritud efektide mõju. Vajaduse korral kasutatakse selleks ML meetodist saadud seost (4.19) (Searle jt., 1992).

Juhuslikule veale vastava dispersioonikomponendi  $\sigma_e^2$  hinnangu saamiseks tuleb võrrandis (4.33) võtta  $i = 0$ . Siis

$$\text{tr}(P) = (N - r(X))\sigma_e^{-2} - \left(\sum_{i=1}^s \hat{u}_i' \hat{u}_i k_i\right) \sigma_e^{-4} \quad (4.34)$$

(Mao, Schaeffer, 1993) ja võrrandi parem pool avaldub valemist (4.15). Võrdsustades seosed (4.15) ja (4.34), saame

$$\hat{\sigma}_e^2 = (y'y - \tilde{b}'X'y - \sum_{i=1}^s \hat{u}_i'Z_i'y) / (N - r(X)). \quad (4.35)$$

Dispersioonikomponentide  $\sigma_i^2$ ,  $i = 1, \dots, s$ , hindamiseks märgime esmalt, et maatriks  $P$  on avaldatav seosega

$$P = S - SZ(Z'SZ + G^{-1})^{-1}Z'S, \quad (4.36)$$

kus  $S = R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}$  (Searle jt., 1992). Esituse (4.36) kuju on sama, mis seosel (4.25), vaid maatriksi  $R^{-1}$  kohal paikneb maatriks  $S$ . Sellest lähtudes on võrduse (4.33) parem pool lahtikirjutatav analoogselt võrduste ahelale ML meetodi algoritmist. Tulemuseks saame võrdused

$$\text{tr}(PZ_i Z_i') = \text{tr}(I_{n_i} - C_{ii} k_i) \sigma_i^{-2} = n_i \sigma_i^{-2} - \sigma_e^2 \text{tr} C_{ii} \sigma_i^{-4}, \quad (4.37)$$

kus  $C = (Z'MZ + G^{-1} \sigma_e^2)^{-1}$ . Võrduse (4.33) parem pool esitub valemiga (4.29).

Võrdsustades seosed (4.37) ja (4.29), saame  $\sigma_i^2, i = 1, \dots, s$ , hinnanguks

$$\hat{\sigma}_i^2 = (\hat{u}_i' \hat{u}_i + \hat{\sigma}_e^2 \text{tr} C_{ii}) / n_i.$$

Nii siin toodud REML kui ka eelmise peatüki ML algoritmi kirjeldus baseeruvad Hendersoni segamudeli võrranditel ja on ühed EM algoritmi (*expectation-maximization*) variantidest. Programmides kasutatakse tavaliselt märksa kiiremini koonduvaid meetodeid, mis põhinevad logaritmilise tõepärafunktsiooni teisel tuletisel. Need meetodid kuuluvad „numbriliste meetodite” valdkonda ning ei pruugi alati saada tulemuseks mittenegatiivselt määratud dispersioonimatrikseid  $R$  või  $G$ . Sõltuvalt programmist ja meetodist asendatakse negatiivsed väärtused tavaliselt väga väikese positiivse väärtusega või nulliga (Searle jt., 1992).

## 5. Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodi võrdlus dispersioonikomponentide ja päritavuskoeffitsiendi hindamisel

Tundes erinevaid meetodeid dispersioonikomponentide hindamiseks, jääb ikkagi vastusetä küsimus, milline neist on parim. ANOVA, Hendersoni 1., 2. ja 3. meetod ning MIVQUE meetod annavad nihketa hinnangud, kuid ei garanteeri nende legaalsust (paiknemist parameeterruumis). ML ja REML meetodil saadud dispersioonikomponentide hinnangud on alati positiivsed, kuid nihkega. Lihtsaim viis erinevate geneetilisi parameetreid hindavate statistiliste meetodite võrdlemiseks on rakendada neid modelleeritud andmetele, mille korral on teada hinnatavate parameetrite tegelikud väärtused (Lin, McAllister, 1984). Antud töös viidi modelleerimised läbi viie dispersioonikomponentide hindamisel enamkasutatava meetodi võrdlemiseks.

Andmed genereeriti vastavalt mudelile (3.1), üldkeskmise  $\mu$  võeti andmete genereerimisel nulliks. Isa mõju genereeriti normaaljaotusega  $N(0,1)$  ja juhuslik viga normaaljaotusega  $N(0,20)$ . Päritavuskoeffitsiendi väärtus on seega 0,1905. Genereeriti 10 isa mõjud, iga isa järglaste arvuks valiti juhuslikult (ühtlase jaotusega) arv 5-st 100-ni. Andmestiku suurus valiti küllalt väike lähtudes arvutuste keerukusest ja kasutada olnud arvuti võimsusest. Näiteks sama tütarde arvu, kuid 25 isa andmete katselisel analüüsil võttis modelleerimise üks ring aega 1,5 tundi.

Hinnangud MIVQUE(0), ML ja REML meetodil leiti MIXED protseduuri (SAS Institute Inc., 1996) abil, Hendersoni 3. meetodi ja MIVQUE hinnangute leidmiseks kirjutati SAS maatriksarvutuse keeles IML (SAS Institute Inc., 1990) programmid. Kuna MIVQUE(0) hinnangud võivad teoreetiliselt tulla negatiivsed, kuid SAS MIXED protseduur asendab negatiivsed hinnangud nulliga, kirjutati ka MIVQUE(0) hinnangute leidmiseks maatriksarvutuse keeles IML programm, mis käivitub vaid juhul, kui MIXED protseduur väljastab nulliga võrduva hinnangu.

Vastupidiselt reaalsele katseandmetele on arvuti abil genereeritud andmete korral isa ja vea mõjud teada ning  $\sigma_{\sigma}^2$ ,  $\sigma_e^2$  ja  $h^2$  hinnatavad ka otse valemeist

$$\sigma_{\sigma}^2 = \frac{\sum(\sigma_i - \bar{\sigma})^2}{n-1}, \quad \sigma_e^2 = \frac{\sum(e_{ij} - \bar{e})^2}{N-1} \quad \text{ja} \quad h^2 = \frac{4\sigma_{\sigma}^2}{\sigma_{\sigma}^2 + \sigma_e^2},$$

kus  $n$  on isade arv ja  $N$  on kõigi tütarde arv andmestikus. Nende otseste hinnangute suhtes on leitavad erinevate meetoditega saadavate hinnangute keskmised ruutvead.

Kokku viidi läbi 500 modelleerimist. Keskmise andmestiku suurus oli 521 rida, keskmine juhuslikust veast tingitud dispersiooni otsene hinnang  $\sigma_e^2 = 20,003$ , keskmine isa mõjust tingitud dispersiooni otsene hinnang  $\sigma_s^2 = 0,987$  ja keskmine päritavuskoefitsiendi otsene hinnang  $h^2 = 0,1889$ .

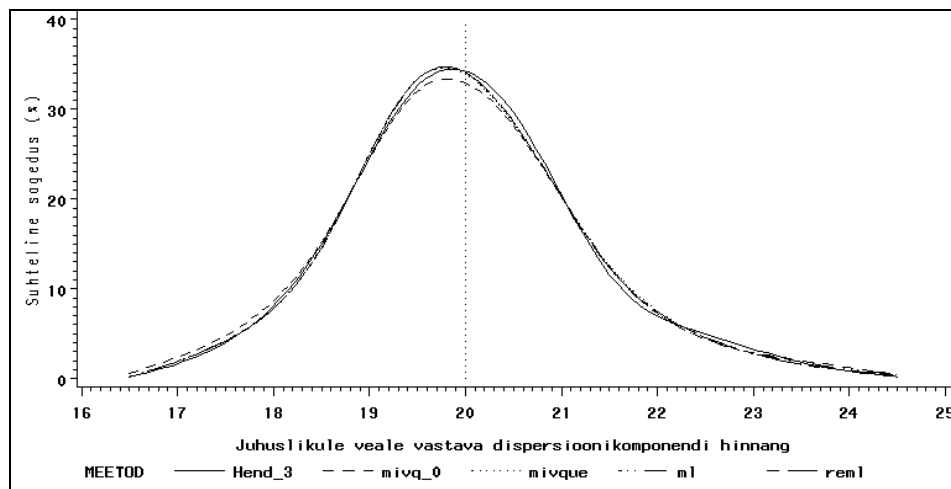
Kõigi meetodite korral arvutati saadud hinnangute keskväärtused, standardhälbed ja keskmised ruutvead. Leiti ka negatiivse või nulliga võrduva päritavuskoefitsiendi hinnangu saamise tõenäosus (kõigil katsetel oli  $h^2$  hinnang väiksem ühest). Tulemused on toodud tabelis 5.1.

Päritavuskoefitsiendi hinnangute osas erineb teistest suurimal määral ML hinnang – see on väikseima keskmise väärtusega ning samas ka väikseima keskmise ruutveaga. Sama saab öelda ka isast tingitud dispersiooni hinnangu kohta. Nulliga võrduvaid päritavuskoefitsiendi hinnanguid esineb ML meetodi korral peaaegu poole rohkem võrreldes teiste meetoditega. Selgelt suurima varieeruvusega hinnangud annab MIVQUE(0) meetod, mis on ka üsna loomulik, sest nulli kasutamine isast tingitud varieeruvuse algjärgendina lihtsustab meetodit tunduvalt, kuid vähendab samal ajal hinnangute täpsust. Hendersoni 3., MIVQUE ja REML meetodil saadud hinnangud erinevad üksteisest vähe. Parameetrite tegelikele väärtustele lähedaseimad keskmised hinnangud on MIVQUE(0) hinnangud, millest palju ei erine ka Hendersoni 3., MIVQUE ja REML meetodil saadud keskmised hinnangud.

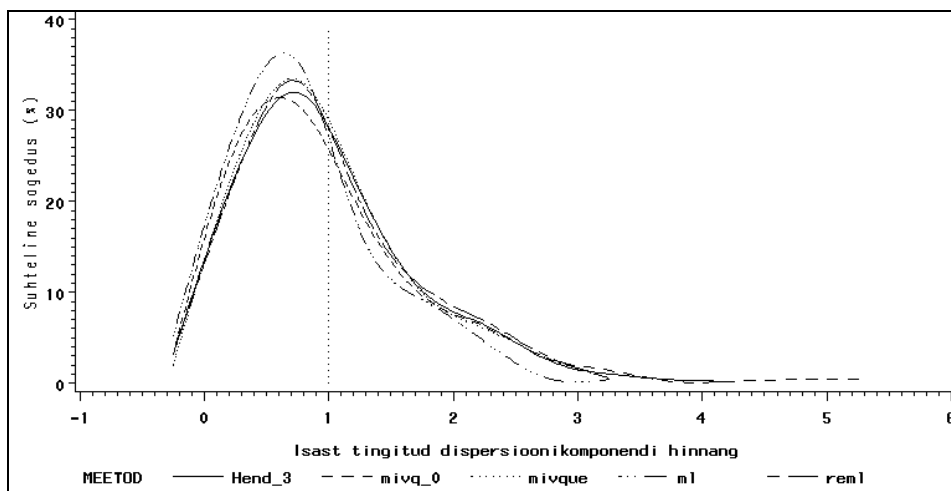
Hinnangute jaotumisest annavad parema pildi joonised 5.1, 5.2 ja 5.3. Juhuslikule veale vastava dispersioonikomponendi hinnangud (joonis 5.1) jaotuvad olenemata hindamismeetodist ühtemoodi, lähedaselt normaaljaotusele, vaevumärgatava parempoolse „sabaga”. ML hinnang langeb teiste hinnangutega kokku, kuna andmed genereeriti vastavalt täiesti juhuslikule mudelile ja üldkeskmise  $\mu$  hindamiseks kulus vaid üks vabadusaste, mille arvesse mittevõtmine hinnangut olulisel määral ei mõjuta. Ka isa mõjust tingitud varieeruvuse hinnangud (joonis 5.2) jaotuvad erinevate meetodite korral sarnaselt. Hendersoni 3. meetodi ja MIVQUE(0) hinnangute jaotus erinevad teistest vaid pikema parempoolse „saba” poolest ning ML hinnangud empiirilise jaotuse teravama tipu poolest. Päritavuskoefitsiendi hinnangute  $\hat{h}^2$  empiiriline jaotus (joonis 5.3) on küllalt sarnane  $\sigma_s^2$  hinnangute jaotusele.

**Tabel 5.1.** 500 modelleerimise tulemused; 10 isa, iga isa tütarde arv ühtlasest jaotusest  $U(5, 100)$ , parameetrite tegelikud väärtused:  $\sigma_e^2 = 20$ ,  $\sigma_\sigma^2 = 1$  ja  $h^2 = 0,1905$ .

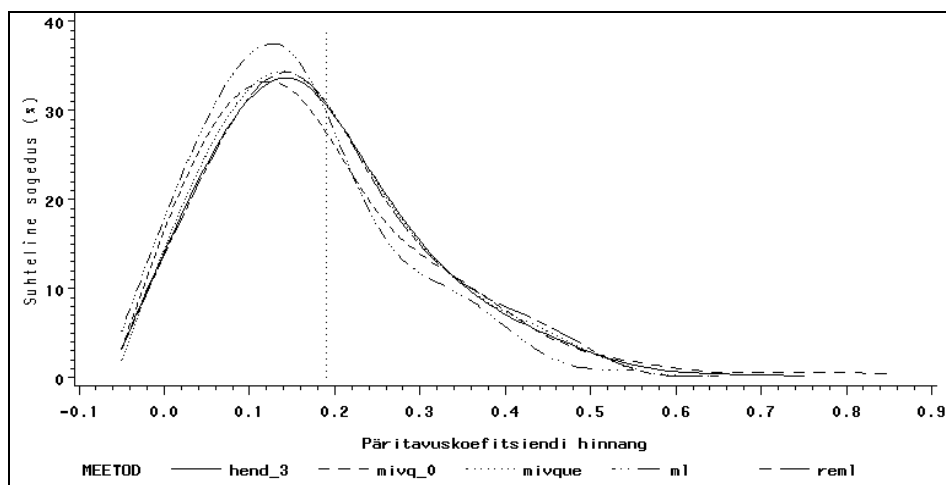
Hinnatav parameeter	Karakteristik	Hindamismeetod				
		Hend. 3.	MIVQUE(0)	MIVQUE	ML	REML
$\sigma_e^2$	keskm.	20,0051	19,9914	20,0038	20,0023	20,0031
	st. hälve	1,2317	1,2687	1,2341	1,2339	1,2342
	keskm. ruutviga	0,0344	0,0706	0,0319	0,03235	0,03255
$\sigma_\sigma^2$	keskm.	0,9915	1,0069	0,9745	0,8338	0,9821
	st. hälve	0,7191	0,8222	0,6778	0,6145	0,6892
	keskm. ruutviga	0,2316	0,3108	0,2304	0,2051	0,2400
$h^2$	keskm.	0,1853	0,1880	0,1826	0,1574	0,1839
	st. hälve	0,1279	0,1459	0,1212	0,1116	0,1231
	keskm. ruutviga	0,0075	0,0098	0,0074	0,0068	0,0077
	$P\{\hat{h}^2 \leq 0\}$	0,032	0,032	0,020	0,052	0,032



**Joonis 5.1.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud  $\sigma_e^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , parameetri tegelikul väärtusel  $\sigma_e^2 = 20$ .



**Joonis 5.2.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud  $\sigma_{\sigma}^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , parameetri tegelikul väärtusel  $\sigma_{\sigma}^2 = 1$ .



**Joonis 5.3.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud päritavuskoefitsiendi  $h^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , päritavuskoefitsiendi tegelikul väärtusel  $h^2 = 0,1905$ .

Kuna kõik võrreldud meetodid on mõeldud dispersioonikomponentide hindamiseks segamudelitest, viidi läbi täiendav modelleerimine, kus mudelisse (3.1) lisati karja fikseeritud mõju. Andmed genereeriti vastavalt mudelile

$$y_{ijk} = \mu + f_i + \sigma_j + e_{ijk}, \quad (5.1)$$

kus  $y_{ijk}$  on isa  $j$  järglane  $k$  karjas  $i$ . Karja efekt  $f_i$  on fikseeritud, isa mõju  $\sigma_j$  on juhuslik,  $\mu$  on üldkeskmine, mis andmete genereerimisel võeti nulliks, ja  $e_{ijk}$  on juhuslik viga.

Uuritava tunnuse koguvarieeruvusest pool loeti tingituks karja (keskkonna) mõjust, mis genereeriti kui normaaljaotusega  $N(0,21)$  juhuslik muutuja, kuid loeti analüüsil fikseerituks. Isa mõju genereeriti normaaljaotusega  $N(0,1)$  ja juhuslik viga normaaljaotusega  $N(0,20)$ . Päritavuskoefitsiendi väärtus on seega 0,1905. Genereeriti 10 isa mõjud, iga isa järglaste arvuks valiti juhuslikult arv 5-st 100-ni. Iga modelleerimise korral jagati isade järglased karjade vahel juhuslikult. Kogu modelleerimist teostava programmi tekst on toodud lisa.

Kokku viidi läbi 500 modelleerimist. Keskmise andmestiku suurus oli 525 rida, keskmine juhuslikust veast tingitud dispersiooni otsene hinnang  $\sigma_e^2 = 19,981$ , keskmine isa mõjust tingitud dispersiooni otsene hinnang  $\sigma_\sigma^2 = 0,989$  ja keskmine päritavuskoefitsiendi otsene hinnang  $h^2 = 0,1893$ .

Kõigi meetodite korral arvutati saadud hinnangute keskväärtused, standardhälbed ja keskmised ruutvead. Leiti ka negatiivse või nulliga võrduva päritavuskoefitsiendi hinnangu saamise tõenäosus. Tulemused on toodud tabelis 5.2.

ML meetodiga leitud jääkvarieeruvuse hinnangu keskmine ruutviga on üle 3 korra suurem kui Hendersoni 3., MIVQUE ja REML meetodiga leitud  $\hat{\sigma}_e^2$  keskmine ruutviga, MIVQUE(0) meetodiga on vahe väiksem, kuid ikkagi üle 2 korra. Siin mängib ilmset rolli fikseeritud mõjude hindamiseks kuluva 10 vabadusastme ignoreerimine, mis tingib  $\sigma_e^2$  alahindamise (vt. joonis 5.4). Samas on aga kõigi parameetrite ML hinnangud väikseima varieeruvusega (vt. ka joonised 5.5 ja 5.6). Ülejäänud meetoditega saadud  $\sigma_\sigma^2$  ja  $h^2$  hinnangud käituvad küllalt sarnaselt, jaotudes tugeva parempoolse ebasümmeetriaga. Võrreldes täiesti juhusliku mudeliga (3.1), ei kahane hinnangute sagedus hinnangu väärtuse kasvades enam nii kiiresti (jaotus on laugem), kuid saavutab miinimumi peaaegu sama kaugel keskmisest väärtusest. Hendersoni 3. ja MIVQUE(0) meetodi korral on „saba” lühem kui täiesti juhuslikul juhul, kuid teiste meetoditega

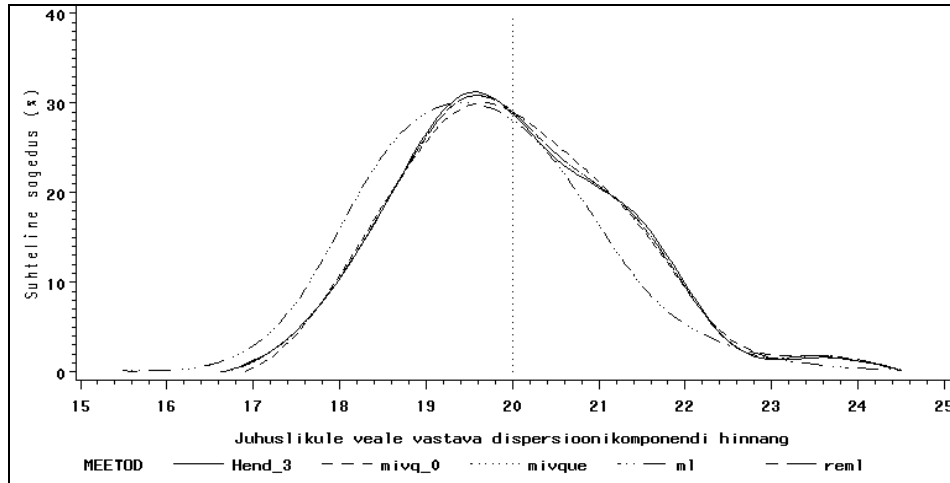


võrreldes siiski pikem. Tingitud on see MIVQUE(0) hinnangu korral meetodi kasutatavast triviaalsest alglähendist ja Hendersoni 3. meetodi korral normaalkõrrandi kasutamisest segamudeli võrrandi asemel (s.t. juhusliku efekti lugemisest fikseerituks). Päritavuskoefitsiendi tegelikule väärtusele lähima hinnangu annab REML meetod, dispersioonikomponentide tegelikele väärtustele on lähimad MIVQUE(0) hinnangud.

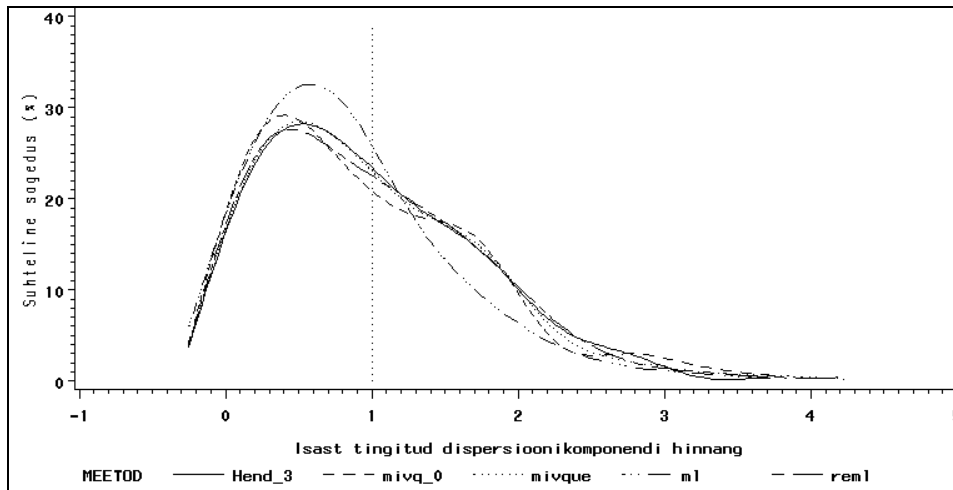
Karjade arvu suurendamine vähendas juhuslikule veale vastava dispersiooni-komponendi  $\sigma_e^2$  ML hinnangu keskmist väärtust ja suurendas keskmist ruutviga. Viies katseks läbi 25 modelleerimist karjade arvu 50 korral, oli  $\sigma_e^2$  ML hinnangu keskmine ruutviga 25 korda suurem ja keskvärtus 1,8 võrra väiksem teiste hinnangutega saadust. Ülejäänud hinnangud käitusid analoogselt tabelis 5.2 toodule.

**Tabel 5.2.** 500 modelleerimise tulemused; andmed genereeritud mudeli (5.1) alusel, 10 isa, iga isa tütarde arv ühtlasest jaotusest  $U(5, 100)$ , parameetrite tegelikud väärtused:  $\sigma_e^2 = 20$ ,  $\sigma_\sigma^2 = 1$  ja  $h^2 = 0,1905$ .

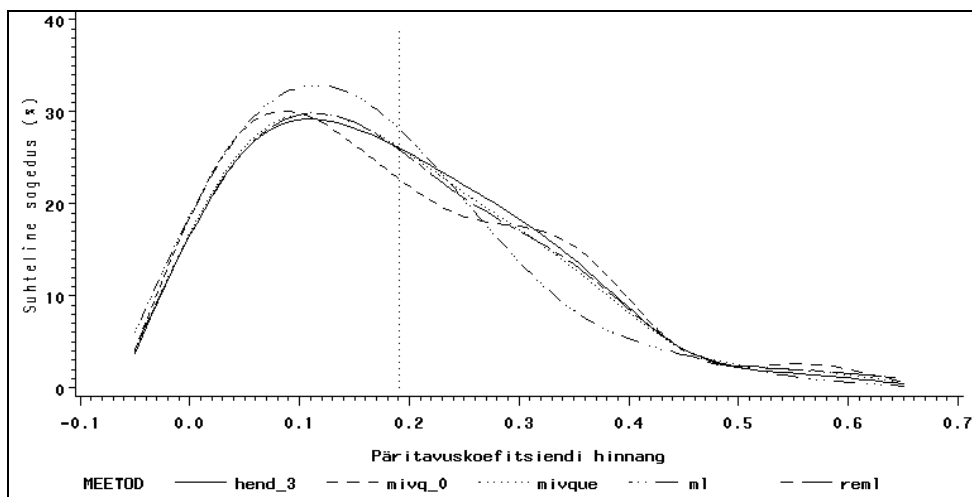
Hinnatav parameeter	Karakteristik	Hindamismeetod				
		Hend. 3.	MIVQUE(0)	MIVQUE	ML	REML
$\sigma_e^2$	keskm.	19,9728	19,9834	19,9757	19,6155	19,9724
	st. hälve	1,2911	1,2957	1,2911	1,2696	1,2909
	keskm. ruutviga	0,0591	0,0898	0,0586	0,1971	0,0593
$\sigma_\sigma^2$	keskm.	1,0118	0,9996	1,0113	0,8695	1,0233
	st. hälve	0,7278	0,7659	0,7405	0,6682	0,7521
	keskm. ruutviga	0,2330	0,2629	0,2629	0,2265	0,2848
$h^2$	keskm.	0,1893	0,1868	0,1891	0,1667	0,1912
	st. hälve	0,1306	0,1375	0,1325	0,1232	0,1340
	keskm. ruutviga	0,0074	0,0085	0,0085	0,0075	0,0088
	$P\{\hat{h}^2 \leq 0\}$	0,038	0,042	0,040	0,060	0,042



**Joonis 5.4.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud  $\sigma_e^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , parameetri tegelikul väärtusel  $\sigma_e^2 = 20$ . Andmed on genereeritud mudeli (5.1) alusel.



**Joonis 5.5.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud  $\sigma_\sigma^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , parameetri tegelikul väärtusel  $\sigma_\sigma^2 = 1$ . Andmed on genereeritud mudeli (5.1) alusel.



**Joonis 5.6.** Hendersoni 3., MIVQUE(0), MIVQUE, ML ja REML meetodiga leitud päritavuskoefitsiendi  $h^2$  hinnangute empiiriline jaotus 10 isa andmetel, iga isa tütarde arvul ühtlasest jaotusest  $U(5, 100)$ , päritavuskoefitsiendi tegelikul väärtusel  $h^2 = 0,1905$ . Andmed on genereeritud mudeli (5.1) alusel.

Hinnangute käitumise uurimiseks suurema andmestiku korral viidi läbi modelleerimised isade arvul 100. Ühe isa tütarde arv valiti endiselt juhuslikult 5-st 100-ni, seega planeeriti andmestiku keskmiseks suuruseks 5250 rida. Andmed genereeriti mudeli (5.1) järgi analoogselt eelmise modelleerimisega. Kuna andmemaatriks oli SAS IML-s tehete tegemiseks liiga suur, leiti parameetrite hinnangud vaid protseduuri MIXED abil, kasutades meetodeid MIVQUE(0), ML ja REML.

Viidi läbi 250 modelleerimist. Keskmine andmemaatriksi ridade arv oli 5218, keskmine juhuslikust veast tingitud dispersioon  $\sigma_e^2 = 20,035$ , keskmine isa mõjust tingitud dispersioon  $\sigma_s^2 = 1,001$  ja keskmine päritavuskoefitsiendi väärtus  $h^2 = 0,1904$ . Hinnangute keskväärtused, standardhälbed ja keskmised ruutvead on toodud tabelis 5.3.

Nagu näha, ei ole juhusliku faktori suure tasemete arvu korral tulemustes olulist erinevust. Kõik kolm meetodit hindavad parameetreid üsna ühesuguse täpsusega. Mingil määral on teistest ebakindlam MIVQUE(0) hinnang, kuid see on tühine erinevus võrreldes meetodi lihtsusega ML ja REML meetodi kõrval. Täpsemaks võib lugeda ML hinnangu. See on ilmselt tingitud fikseeritud faktori küllalt väikesest tasemete arvust võrreldes juhusliku faktori tasemete arvuga.

**Tabel 5.3.** 250 modelleerimise tulemused; andmed genereeritud mudeli (5.1) alusel, 100 isa, iga isa tütarde arv ühtlasest jaotusest  $U(5, 100)$ , parameetrite tegelikud väärtused:  $\sigma_e^2 = 20$ ,  $\sigma_\sigma^2 = 1$  ja  $h^2 = 0,1905$ .

Hinnatav parameeter	Karakteristik	Hindamismeetod		
		MIVQUE(0)	ML	REML
$\sigma_e^2$	keskm.	20,035	19,996	20,032
	st. hälve	0,380	0,376	0,377
	keskm. ruutviga	0,0072	0,0046	0,0031
$\sigma_\sigma^2$	keskm.	1,020	1,012	1,027
	st. hälve	0,239	0,213	0,215
	keskm. ruutviga	0,0228	0,0212	0,0223
$h^2$	keskm.	0,193	0,192	0,195
	st. hälve	0,043	0,038	0,039
	keskm. ruutviga	0,00076	0,00070	0,00073

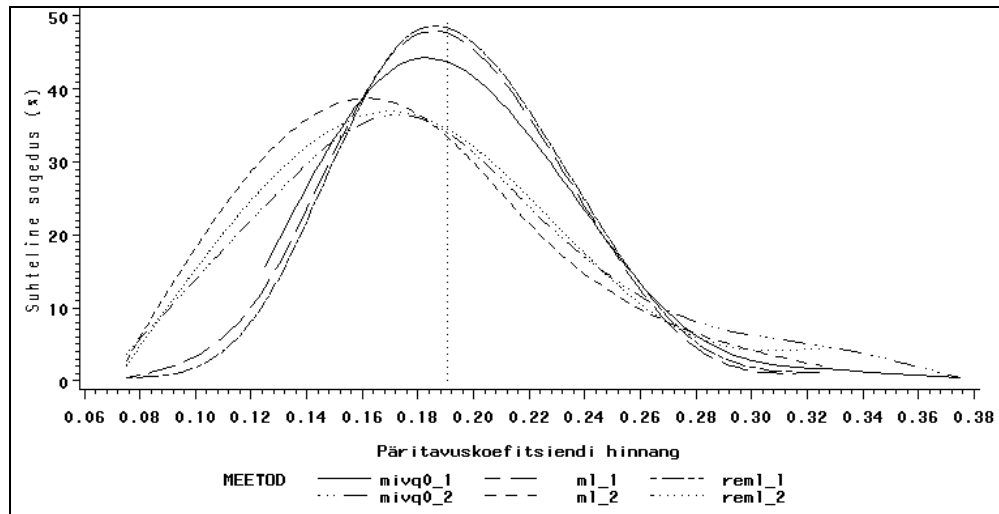
Saamaks ettekujutust hinnangute käitumisest sama andmestiku suuruse, kuid erineva isade ja tütarde arvu korral, viidi läbi modelleerimised, mille käigus genereeriti andmed vastavalt 25 isale ja tütarde arvule 162-st 258-ni (et saada andmestiku planeeritud suuruseks endiselt 5250 rida). Viidi läbi 250 modelleerimist. Keskmise andmemaatriksi ridade arv oli 5239, keskmine juhuslikust veast tingitud dispersioon  $\sigma_e^2 = 20,011$ , keskmine isa mõjust tingitud dispersioon  $\sigma_\sigma^2 = 0,985$  ja keskmine päritavuskoefitsiendi väärtus  $h^2 = 0,1880$ . Modelleerimise tulemused on toodud tabelis 5.4.

Isade arvu vähendamine tõi kaasa ML hinnangute keskmiste ruutvigade suurenemise ja hinnangute keskmiste väärtuste vähenemise võrreldes MIVQUE(0) ja REML hinnangutega. MIVQUE(0) ja REML meetodite vahelised erinevused on sama minimaalsed kui eelmise modelleerimise korral. Hinnangute käitumine erineva isade ja tütarde arvu korral tuleb selgeimalt esile joonisel 5.7, kus on kujutatud päritavuskoefitsiendi hinnangute jaotus MIVQUE(0), ML ja REML meetodi ning isade arvu 100 ja 25 (andmestiku suurus 5250 rida) korral. Isast tingitud varieeruvuse alusel

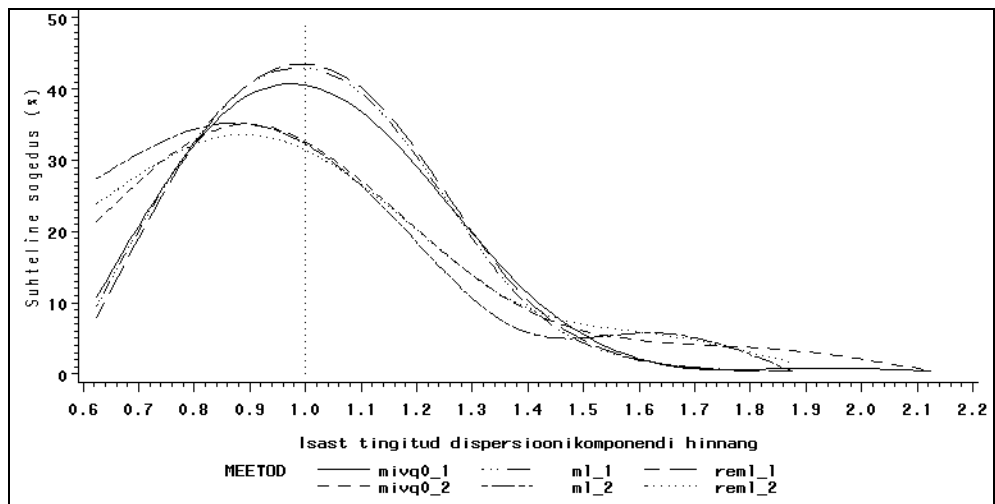
leitud päritavuskoefitsiendi hinnangu täpsus suureneb isade arvu kasvades tunduvalt kiiremini kui tütarde arvu kasvades. Väiksema isade arvu korral on  $h^2$  hinnangu jaotus raske parempoolse „sabaga” ning seda isegi hoolimata keskmiselt 220 tütrest isa kohta. Samasugune on olukord ka isast tingitud dispersiooni hindamisel (joonis 5.8). Juhuslikule veale vastav dispersioonikomponent ei ole isade arvu langusele eriti tundlik, kuid tabelite 5.3 ja 5.4 võrdlemisel hakkab silma, et suurema isade arvu korral on hinnangud veidi täpsemad (vt. ka joonis 5.9).

**Tabel 5.4.** 250 modelleerimise tulemused; andmed genereeritud mudeli (5.1) alusel, 25 isa, iga isa tütarde arv ühtlasest jaotusest  $U(162; 258)$ , parameetrite tegelikud väärtused:  $\sigma_e^2 = 20$ ,  $\sigma_\sigma^2 = 1$  ja  $h^2 = 0,1905$ .

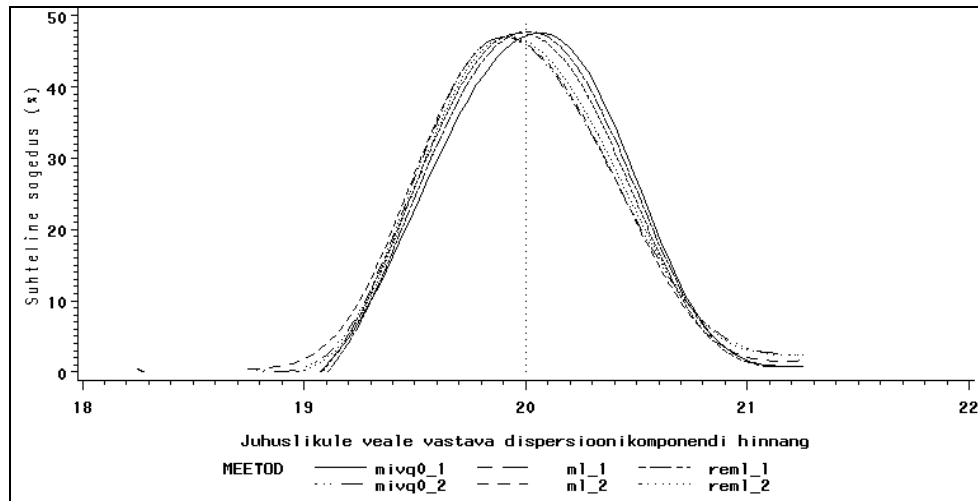
Hinnatav Parameeter	Karakteristik	Hindamismeetod		
		MIVQUE(0)	ML	REML
$\sigma_e^2$	keskm.	20,014	19,981	20,015
	st. hälve	0,416	0,412	0,412
	keskm. ruutviga	0,0023	0,0018	0,0009
$\sigma_\sigma^2$	keskm.	0,981	0,934	0,977
	st. hälve	0,319	0,286	0,298
	keskm. ruutviga	0,0171	0,0177	0,0161
$h^2$	keskm.	0,186	0,178	0,186
	st. hälve	0,057	0,052	0,054
	keskm. ruutviga	0,00057	0,00058	0,00054



**Joonis 5.7.** Päritavuskoefitsiendi  $h^2$  MIVQUE(0), ML ja REML hinnangute empiiriline jaotus, 100 isa ja tütarde arv 5-100 (mivq0\_1, ml\_1 ja reml\_1) ning 25 isa ja tütarde arv 162-258 (mivq0\_2, ml\_2 ja reml\_2), tegelik  $h^2 = 0,1905$ .



**Joonis 5.8.** Isa mõjule vastava dispersioonikomponendi  $\sigma^2_{\sigma}$  MIVQUE(0), ML ja REML hinnangute empiiriline jaotus; 100 isa ja tütarde arv 5-100 (mivq0\_1, ml\_1 ja reml\_1) ning 25 isa ja tütarde arv 162-258 (mivq0\_2, ml\_2 ja reml\_2), tegelik  $\sigma^2_{\sigma} = 1$ .



**Joonis 5.9.** Juhuslikule veale vastava dispersioonikomponendi  $\sigma_e^2$  MIVQUE(0), ML ja REML hinnangute empiiriline jaotus; 100 isa ja tütarde arv 5-100 (mivq0\_1, ml\_1 ja reml\_1) ning 25 isa ja tütarde arv 162-258 (mivq0\_2, ml\_2 ja reml\_2), tegelik  $\sigma_e^2 = 20$ .

MIVQUE(0), ML ja REML hinnanguid on uuritud ka ekstremaalsete parameetriväärtuste korral ( $h^2 = 0,01$  ja  $h^2 = 1$ ). Selleks viidi mõlemal juhul läbi 25 modelleerimist tabeli 5.3 pealkirjas toodud tingimustel. Tulemuste alusel võib öelda, et päritavuskoeffitsiendi väärtuse lähenemisel nullile selle hindamise täpsus väheneb, suhte  $\sigma_\alpha^2/\sigma_e^2$  lähenemisel ühele kolmandikule ( $h^2 \rightarrow 1$ ) aga hindamise täpsus suureneb. Meetoditevahelised erinevused jäid parameetri väärtuste muutudes samasugusteks nagu tabelis 5.3.

## Kokkuvõte

Töö põhitulemuseks on mõningate loomakasvatuses laialt kasutatavate statistiliste meetodite kriitiline analüüs. Selleks on kirjanduse põhjal refereeritud ja ka iseseisvalt tuletatud mitmed teoreetilised seosed päritavuskoeffitsiendi ja geneetiliste dispersiooni-komponentide hinnangute kohta ning läbi viidud hulk modelleerimiskatseid.

Tasakaalulise andmestiku analüüsil selgus, et nii dispersioonikomponendid kui ka nende hinnangute dispersioonid on dispersioonanalüüsi abil nihketa hinnatavad. Päritavuskoeffitsiendi hinnang ja selle dispersiooni hinnang on aga modelleerimistulemuste alusel ilmselt nihkega hinnangud. Seejuures on  $h^2$  hinnangud tegelikust väärtusest pisut väiksemad, dispersiooni hinnang aga tegelikust dispersioonist suurem.

Päritavuskoeffitsiendi illegaalse hinnangu tõenäosus pole piisavalt suure isade ja tütarde arvu korral kuigi suur. Probleeme võib tekkida vaid nulli- või ühelähedaste  $h^2$  väärtuste korral. Väga väikese päritavuse korral kahaneb hinnangu illegaalsuse tõenäosus kiiremini tütarde arvu kasvades isa kohta, suure päritavuse korral aga isade arvu kasvades.

Neljandas peatükis on esitatud ühtse ülesehitusega algoritmid olulisemate dispersioonikomponente mittetasakaalulisel juhul hindavate meetodite kohta. Jääkvarieeruvuse hindamisel on kohati piiratud pikemate tõestusteta tulemuste äratoomisega. Kõik esitatud algoritmid on kontrollitud. Selleks rakendati katseandmetele SAS-paketi VARCOMP protseduuri (sisaldab Hendersoni 3., MIVQUE(0), ML ja REML meetodit) ning SAS maatriksarvutuse keeles IML realiseeritud algoritme. Samade meetoditega saadud hinnangud olid identsed. MIVQUE algoritmi rakendati kontrolliks Hendersoni (1984) ja Mao ning Schaeffer'i (1993) näidetele, saadud hinnangud olid identsed raamatuis toodutega.

Mittetasakaalulisel juhul kasutatavate hindamismeetodite võrdlemisel modelleerimise abil selgus, et suure fikseeritud efektide tasemete arvu korral pole soovitatav kasutada ML meetodit, kuna see hindab dispersioonikomponendid tegelikust väiksemaks. MIVQUE meetod, mis arvutustehniliselt on kõige keerukam (nõuab segamudeli võrrandi terve koeffitsientide maatriksi pööramist), ei oma märkimisväärset eelist dispersioonikomponentide hindamisel ja seega pole tema kasutamine ressursside kokkuhoiu huvides mõttekas. MIVQUE(0) hinnangud olid küll kõige ebasümmeetri-



lisemad (parempoolse „sabaga”), kuid hinnangute varieeruvus ja keskväärtus ei erinenud teistest meetoditest kuigivõrd. Seetõttu on suurema andmestiku korral oma lihtsuse tõttu eelistatud MIVQUE(0) meetod. Hendersoni 3. meetod andis töös vaadeldud lihtsate mudelite puhul samuti küllaltki head tulemused. Arvestades, et loomakasvatases vaadeldakse juhuslikuna tavaliselt vaid üht faktorit - isa, ei saa ka Hendersoni 3. meetodit „maha kanda”. Kõige universaalsem tundub olevat REML meetod. Seda arvamust kinnitab ka REML meetodi laialdane rakendamine praktikas.

Meetodite paremusjärjestusse seadmisest olulisemgi (loomakasvatuse jaoks) on tulemus, et isa mõjust tingitud dispersioonikomponendi abil avalduva päritavuskoefitsiendi hinnangu täpsus sõltub isade arvust märksa enam, kui tütarde arvust isa kohta. See väide seab kahtluse alla aretuses laialt levinud seisukoha, et geneetiliste parameetrite täpsemaks hindamiseks on vaja eelkõige suurendada tütarde arvu isa kohta. Muidugi vajab viimane tulemus täiendavat kontrollimist.

Samuti jäävad tulevikku uuringud küsimuses, kuidas muudab geneetiliste parameetrite hinnangut loomadevaheliste sugulussidemete arvestamine. Siin on kindlasti abiks vastavate programmide olemasolu, sest loomade arvuga võrdse dimensiooniga maatriksid on SAS IML-s tehete tegemiseks oletatavasti liiga suured. Parandada tuleks ka koostööd Vabariikliku Jõudluskontrollikeskusega, kus seda, nn. looma mudelit, rakendatakse praktilistes arvutustes.

## **Kasutatud kirjandus**

1. Becker, W. A. Manual of Quantitative Genetics. Pullman, Washington: Academic Enterprises, 1984
2. Boldman, K. G., Kriese, L. A., Van Vleck, L. D., Van Tassell, C. P., Kachman, S. D. A Manual for Use of MTDFREML. A Set of Programs To Obtain Estimates of Variances and Covariances (DRAFT). U.S. Department of Agriculture, Agricultural Research Service, 1995
3. Brascamp, P., Bovenhuis, H., Van der Werf, J. Animal genetics. Lecture notes. Wageningen Agricultural University, 1995
4. Falconer, D. S. Introduction to quantitative genetics. New York: Longman, 1981
5. Groeneveld, E. PEST User's Manual. Institute of Animal Husbandry and Animal Behaviour, Federal Agricultural Research Centre (FAL), Germany, 1995
6. Harvey, W. R. User's Guide for LSMLMW and MIXMDL PC-2 version. Mixed model least-squares and maximum likelihood computer program, 1990
7. Harville, D., A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistician Association, 72: 320-340, 1977
8. Henderson, C. R. Applications of Linear Models in Animal Breeding. University of Guelph, 1984
9. Kaart, T. Veiste piimajõudluse statistiline analüüs. Diplomitöö. Tartu Ülikool, 1995
10. Kaart, T. Statistical analysis of heritability coefficient. Proceedings of the 2<sup>nd</sup> Baltic Animal Breeding Conference: 25-27, Kaunas, 1996
11. Kaart, T. Probability of the estimation of heritability being negative or greater than one. Proceedings of the 3<sup>rd</sup> Baltic Animal Breeding Conference: 57-59, Riia, 1997
12. Lin, C. Y., McAllister, A. J. Monte Carlo Comparison of Four Methods for Estimation of Genetic Parameters in the Univariate Case. Journal of Dairy Science, 67: 2389-2398, 1984
13. Mao, I. L., Schaeffer, L. R. Techniques for Analyzing Unbalanced Data (with emphasis on Applications to Breeding and Genetics Research). Notes for Linear Models and Estimation of Breeding Values. An internordic post-graduate course. Michigan State University, 1993

14. Parring, A.-M., Möls, T. Use of absorption in environmental studies. TÜ toimetised, 968 (Stochastic procedures): 29-40. Tartu Ülikool, 1994
15. Rao, C. R. Estimation of Variance and Covariance Components in Linear Models. Journal of the American Statistical Association, 67: 112-115, 1972
16. SAS Institute Inc. SAS/IML Software: Usage and Reference, Version 6, First Edition. Cary, NC: SAS Institute Inc., 1989
17. SAS Institute Inc. SAS/STAT Software: Changes and Enhancements through Release 6.11. Cary, NC: SAS Institute Inc., 1996
18. Searle, S. R. Linear Models for Unbalanced Data. New York etc.: John Wiley & Sons, 1987
19. Searle, S. R., Casella, G., McCulloch, C. E. Variance Components. New York etc.: John Wiley & Sons, 1992
20. Teinberg, R. Põllumajandusloomade geneetika. Tallinn: Valgus, 1978
21. Tiit, E., Parring, A., Möls, T. Tõenäosusteooria ja matemaatiline statistika. Tallinn: Valgus, 1977
22. Van Arendonk, J., Van der Werf, J., Groen, A. Breeding value estimation. Lecture notes for E250-209. Wageningen Agricultural University, 1997

# **Estimation of variance components and heritability coefficient in animal populations**

**T. Kaart**

## **Abstract**

In this master thesis the estimation of most essential genetic parameters as variance components and heritability coefficient  $h^2$  is studied. As usually in animal breeding, the variance component caused by sire is analysed.

For convenience, basic concepts and useful notations in breeding theory (chapter one) and general linear models theory (chapter two) are presented.

The third chapter describes the estimation of variance components and heritability coefficient in balanced designs. The 1-way random model is used. From the computer simulation we may conclude that the estimates of variance components are unbiased but the estimate of heritability coefficient is biased downward.

In the third chapter, the probability  $P_I$  of an illegal  $h^2$  estimate is calculated. The illegality of the estimate means that  $\hat{h}^2$  is smaller than zero or greater than one. The calculations showed, as would be expected, that  $P_I$  decreases when either the number of sires or number of daughter per sire increases. In general, if the number of sires or the number of daughters per sire are not too small, there seems to be no reason to worry about the possibility of having an illegal estimate of heritability coefficient. The difficulties may occur only if the true  $h^2$  is very close to zero or one.

In the fourth chapter the algorithms for five methods of computing variance components for unbalanced data are described. These methods are Henderson's method 3, MIVQUE, MIVQUE(0), ML and REML.

In the fifth chapter, these five methods for estimation of variance components and heritability coefficient are compared in an extensive computer simulation experiment. The program generates replicates of the 1-way random model and 2-way mixed model data sets and calculates estimates with all five methods. The results demonstrated that Henderson's method 3 and MIVQUE(0)-method are the most unstable. This is caused by treating random effect of sire as fixed (for Henderson's method 3) and using zeros for prior values of variance components (for MIVQUE(0)). Maximum likelihood estimates

were biased downward because ML-method does not take into account the degrees of freedom related with fixed effects. In spite of own complicated computation the MIVQUE-method don't give better estimates than other methods. The most recommended method seems to be REML.

## Lisa. Programm dispersioonikomponentide ja päritavuskoeffitsiendi hindamise modelleerimiseks mittetasakaalulisel juhul segamudeli korral

```
%macro mod(kordusi, isasid, karju);
%local u;
%do u=1 %to &kordusi;                               /* Modelleerimise kordamine 'kordusi' korda */

data andmed;                                         /* Andmete genereerimine */
array karjad[1:10] kar1-kar10;
do kari=1 to &karju;
  karjad[kari]=sqrt(21)*rannor(0);                   /* Karja mõju genereerimine */
end;

do isa=1 to &isasid;
  i=rannor(0);                                       /* Isa mõju genereerimine */
  lehmi=int(5+95*rannor(0));                         /* Isa tütarde arvu määramine */
  do j=1 to lehmi;
    y=0;
    e=sqrt(20)*rannor(0);                            /* Juhusliku mõju genereerimine */
    k_nr=int(1+&karju*rannor(0));                    /* Karja valimine */

    y=e+i+karjad[k_nr];                             /* Uuritavale tunnusele väärtuse omistamine */

  output;
  drop kar1-kar10 kari;
end;
end;                                                 /* Andmete genereerimise lõpp */
dm 'output; clear';

proc means data=andmed var;                          /* Juhuslikule veale vastava dispersioonikomponendi */
var e;                                              /* otsese hinnangu arvutamine */
output out=e_stat var=disp_e;
run;

proc means data=andmed mean; /* Andmestiku, kus iga isa efekt esineb vaid ühel real, moodustamine */
var i;
by isa;
output out=d_isa mean=isa_mqju;
run;

proc means data=d_isa var;                          /* Isa mõjust tingitud dispersioonikomponendi otsene hindamine */
var isa_mqju;
output out=i_stat var=disp_i;
run;

proc mixed data=andmed method=mivque0;              /* MIVQUE(0) hinnangute arvutamine */
class isa k_nr;
model y = k_nr;
random isa;
make 'covparms' out=miv_out;
run;

proc mixed data=andmed method=reml;                  /* REML hinnangute arvutamine */
class isa k_nr;
model y = k_nr;
random isa;
```

```

make 'covparms' out=reml_out;
run;

proc mixed data=andmed method=ml;                                /* ML hinnangute arvutamine */
class isa k_nr;
model y = k_nr;
random isa;
make 'covparms' out=ml_out;
run;

proc iml;                                                        /* Maatriksarvutuse algus */
reset nolog;

use andmed;                                                    /* Andmemaatriksi sisselugemine */
read all into andmed;
y=andmed[, 5];
isa=andmed[, 1];
k_nr=andmed[, 7];

Z=design(isa);                                                  /* Juhuslikele efektidele vastava plaanimaatriksi koostamine */
X=J(nrow(y),1,1)||design(k_nr);                                /* Fikseeritud efektidele vastava plaanimaatriksi koostamine */
W=X||Z;

N=nrow(y);
q=ncol(Z);
p=ncol(X);

gXXX=ginv(X`*X)*X;                                             /* Korduvalt kasutatavad seosed on igal modelleerimise sammul*/
                                                                /* välja arvatud vaid üks kord */
M=I(N)-X*gXXX;

ZMZ=Z`*M*Z;
ZMy=Z`*M*y;

s=ginv(ZMZ)*ZMy;
b=gXXX*y;                                                      /* Fikseeritud efektide vähimruutude hinnangu leidmine */

/* Hendersoni 3. meetod */

viga_hen=(y`*M*y-s`*(ZMy))/(N-round(trace(ginv(W)*W))); /* Dispersioonikomponentide hindamine */
isa_hen=(s`*(ZMy)-viga_hen*(q-1))/(trace(ZMZ));             /* vastavalt valemitele (4.7) ja (4.10) */
h2_hen=4*isa_hen/(isa_hen+viga_hen);

hen_out=viga_hen||isa_hen||h2_hen;                             /* Väljundandmestiku koostamine */
create hen_out from hen_out[colname={viga_hen isa_hen h2_hen}];
append from hen_out;
close hen_out;

/* MIVQUE(0) */

use miv_out;
read all into mivque0;

if mivque0[1,1]=0 then do;                                     /* Kontroll, kas MIXED protseduuri tulemus on null */
  Abi0=J(2,2,0);                                             /* Dispersioonikomponentide hindamine võrdusest (4.18) */
  Abi0[1,1]=trace(ZMZ*ZMZ);
  Abi0[1,2]=trace(ZMZ);

```

```

Abi0[2,1]=Abi0[1,2];
Abi0[2,2]=trace(M);

Q0=J(2,1,0);
Q0[1,1]=ZMy`*ZMy;
Q0[2,1]=y`*y-b`*X`*y;

solution=inv(Abi0)*Q0;
viga_miv=solution[2,];
isa_miv=solution[1,];
h2_miv=4*isa_miv/(isa_miv+viga_miv);

miv2_out=J(2,1,0);
miv2_out[1,1]=isa_miv;
miv2_out[2,1]=viga_miv;
end;
else miv2_out=J(2,1,0);

create miv2_out from miv2_out[colname={est2}];          /* Väljundandmestiku koostamine */

append from miv2_out;
close miv2_out;

/* MIVQUE */

k=viga_hen/isa_hen;
r=W`*y;
nWW=p+q;

B0=W`*W;
B1=W`*Z*Z`*W;
WWk=B0+(J(nWW,p,0))/(J(p,q,0)/(I(q)*k));          /* Segamudeli võrrandi kordajate maatriksi leidmine */

C=ginv(WWk);          /* Segamudeli võrrandi kordajate maatriksi üldistatud pöördmaatriksi leidmine */
C1=(C[,p+1:ncol(C)]);
C1C1=C1`*C1;

Abi=J(2,2,0);
Abi[1,1]=trace(C1C1*B1);
Abi[1,2]=trace(C1C1*B0);
Abi[2,1]=N-trace(C*B1);
Abi[2,2]=N-trace(C*B0);

Q=J(2,1,0);
Q[1,1]=r`*C1C1*r;
Q[2,1]=y`*y-r`*C*r;

Solution=inv(Abi)*Q;          /* Dispersioonikomponentide hindamine võrdusest (4.17) */

vig_mivq=solution[2,];
isa_mivq=solution[1,];
h2_mivq=4*isa_mivq/(isa_mivq+vig_mivq);

mivq_out=vig_mivq||isa_mivq||h2_mivq;          /* Väljundandmestiku koostamine */
create mivq_out from mivq_out[colname={vig_mivq isa_mivq h2_mivq}];
append from mivq_out;
close mivq_out;

quit;          /* Maatriksarvutuse lõpp */

```



```

data miv0_out;                                /* MIVQUE(0)-hinnangu valimine sõltuvalt proc mixed tulemusest */
merge miv_out miv2_out;
if est=0 then est=est2;
run;

data hen_out;                                /* Modelleerimistulemuste ühendamine */
set hen_out;                                  /* (Hendersoni 3. meetodil saadud hinnangud) */
k=&u;
%if &u=1 %then %do;
  data hen;
  set hen_out;
%end;
%else %do;
  data hen;
  set hen hen_out;
%end;

data miv0_out;                                /* Modelleerimistulemuste ühendamine */
set miv0_out;                                  /* (MIVQUE(0) hinnangud) */
k=&u;
%if &u=1 %then %do;
  data miv0;
  set miv0_out;
%end;
%else %do;
  data miv0;
  set miv0 miv0_out;
%end;

data mivq_out;                                /* Modelleerimistulemuste ühendamine */
set mivq_out;                                  /* (MIVQUE hinnangud) */
k=&u;
%if &u=1 %then %do;
  data mivque;
  set mivq_out;
%end;
%else %do;
  data mivque;
  set mivque mivq_out;
%end;

data ml_out;                                  /* Modelleerimistulemuste ühendamine */
set ml_out;                                    /* (ML hinnangud) */
k=&u;
%if &u=1 %then %do;
  data ml;
  set ml_out;
%end;
%else %do;
  data ml;
  set ml ml_out;
%end;

data reml_out;                                /* Modelleerimistulemuste ühendamine */
set reml_out;                                  /* (REML hinnangud) */
k=&u;
%if &u=1 %then %do;
  data reml;
  set reml_out;
%end;

```

```

%else %do;
  data reml;
  set reml reml_out;
%end;

data e_stat;
set e_stat;
k=&u;
%if &u=1 %then %do;
  data e_stat1;
  set e_stat;
%end;
%else %do;
  data e_stat1;
  set e_stat1 e_stat;
%end;

data i_stat;
set i_stat;
k=&u;
%if &u=1 %then %do;
  data i_stat1;
  set i_stat;
%end;
%else %do;
  data i_stat1;
  set i_stat1 i_stat;
%end;

%end;
%mend;

%mod(500, 10, 10);
/* Viiakse läbi 500 modelleerimist, igal sammul genereeritakse 10 isa ja 10 karja andmed */

proc sort data=miv0;
by k covparm;
data mivque0;
set miv0;
by k;
select (covparm);
when ('Residual') do;
  viga_miv=est;
  retain viga_miv;
end;
when ('ISA') do;
  isa_miv=est;
  retain isa_miv;
end;
end;
keep viga_miv isa_miv h2_miv;
if last.k then do;
  h2_miv=4*isa_miv/(isa_miv+viga_miv);
  output;
end;

```

```

proc sort data=ml;                                /* (ML meetod) */
  by k covparm;
  data ml_lqpp;
  set ml;
  by k;
  select (covparm);
  when ('Residual') do;
    viga_ml=est;
    retain viga_ml;
  end;
  when ('ISA') do;
    isa_ml=est;
    retain isa_ml;
  end;
end;
keep viga_ml isa_ml h2_ml;
if last.k then do;
  h2_ml=4*isa_ml/(isa_ml+viga_ml);
  output;
end;

proc sort data=reml;                              /* (REML meetod) */
  by k covparm;
  data reml_lp;
  set reml;
  by k;
  select (covparm);
  when ('Residual') do;
    viga_reml=est;
    retain viga_reml;
  end;
  when ('ISA') do;
    isa_reml=est;
    retain isa_reml;
  end;
end;
keep viga_reml isa_reml h2_reml;
if last.k then do;
  h2_reml=4*isa_reml/(isa_reml+viga_reml);
  output;
end;
run;

data result_1;                                    /* Tulemuste ühendamine ühte faili */
  merge e_stat1 i_stat1 hen mivque mivque0 reml_lp ml_lqpp;
run;

data bubu.sm_mod;
  set result_1;
  h2=4*disp_i/(disp_i+disp_e);                    /* Päritavuskoeffitsiendi otsese hinnangu leidmine */

  MSE_e_he=(disp_e-viga_hen)**2;                 /* Hinnangute ruutvigade leidmine */
  MSE_i_he=(disp_i-isa_hen)**2;
  MSE_h_he=(h2-h2_hen)**2;

  MSE_e_mi=(disp_e-vig_mivq)**2;
  MSE_i_mi=(disp_i-isa_mivq)**2;
  MSE_h_mi=(h2-h2_mivq)**2;

```

```

MSE_e_m0=(disp_e-viga_miv)**2;
MSE_i_m0=(disp_i-isa_miv)**2;
MSE_h_m0=(h2-h2_miv)**2;

MSE_e_ml=(disp_e-viga_ml)**2;
MSE_i_ml=(disp_i-isa_ml)**2;
MSE_h_ml=(h2-h2_ml)**2;

MSE_e_re=(disp_e-viga_re)**2;
MSE_i_re=(disp_i-isa_reml)**2;
MSE_h_re=(h2-h2_reml)**2;
run;

proc means data=bubu.sm_mod;                                /* Hinnangute olulisemate arvarakteristikute leidmine */
var disp_e disp_i h2
    viga_hen isa_hen h2_hen
    viga_miv isa_miv h2_miv
    vig_mivq isa_mivq h2_mivq
    viga_ml isa_ml h2_ml
    viga_reml isa_reml h2_reml;
run;

proc means data=bubu.sm_mod mean;                            /* Hinnangute keskmiste ruutvigade leidmine */
var MSE_e_he MSE_e_mi MSE_e_m0 MSE_e_ml MSE_e_re
    MSE_i_he MSE_i_mi MSE_i_m0 MSE_i_ml MSE_i_re
    MSE_h_he MSE_h_mi MSE_h_m0 MSE_h_ml MSE_h_re;
run;

```