

VII

ÜLDINE LINEAARNE MUDEL LOOMADE ARETUSVÄÄRTUSTE JA POPULATSIOONI GENEETILISTE PARAMEETRITE HINDAMISEL

Eelmises peatükis käsitletud üldiste lineaarsete mudelite parameetrite hindamine moodustab vaid ühe osa tegevustest, mis on vajalikud nende mudelite reaalseks rakendamiseks populatsioonigeneetikas. Sageli on arvutuste korrektsest teostamisest märksa olulisem oskus mõista hinnatud parameetrite sisu ja konkreetse mudeli püstitamisega tehtud eeldusi. See tähendab, et soovides matemaatilise mudeliga tegelikkust võimalikult hästi modelleerida,

- tuleb esimese asjana formuleerida huvi pakkuv tegelikkus nõ geneetilise mudeli kujul (tuleb selgeks mõelda, kuidas võib uuritav tunnus olla geneetiliselt determineeritud);
- seejärel tuleb geneetiline mudel panna ümber matemaatilise statistika keelde (esitada näiteks üldise lineaarse segamudelina), hindamaks selle parameetreid statistika protseduuride abil;
- viimase asjana tuleb sobitada statistilisest mudelist hinnatud parameetrid ja/või nende funktsioonid geneetilisse mudelisse, andmaks neile tegelikkuse suhtes õiget tähendust.

7.1 POOL- JA TÄISÕVEDE MUDELID

7.1.1 Isa mudeli esitus

Pikki aastaid oli valdav meetod geneetiliste parameetrite hindamisel poolõvede analüüs. Selle käigus jagatakse uuritavad loomad gruppidesse vaid ühe vanema järgi, teise vanema geneetiline mõju jäetakse arvesse võtmata (loetakse tühiseks). Et majanduslikult otstarbekam on valida isasloomi (heade geenidega isalt on võimalik saada märksa enam järglasi võrreldes geneetiliselt potentsiaalt sama hea emaga), on selleks vaatlusaluseks vanemaks enamasti isa – siit ka meetodi enamtuntud nimetus, **isa mudel** (*sire model*). Isa mudel hindab geneetilise efektina isa mõju, eeldades vaikimisi, et kõik isad on paaritatud populatsiooni keskmiste emadega ning emad pole omavahel sugulased.

- Geneetika mõistes (nn **geneetilise mudelina**) avaldub isa j järglase k fenotüübiväärtus P_{jk} kujul (vt pt 3.1.1, valem (3.3))

$$P_{jk} = \bar{P} + A_{jk} + E_{jk}$$

mis on isa mudeli kontekstis (vt valem (3.4)) esitatav võrrandina

$$P_{jk} = \bar{P} + \frac{1}{2}A_{S_j} + \underbrace{\frac{1}{2}A_{D_{jk}} + MS_{jk} + E_{jk}}_{E_{jk}} = \bar{P} + \frac{1}{2}A_{S_j} + E_{jk}^* , \quad (7.1)$$

kus

MS_{jk} ja E_{jk} on vastavalt Mendeli valiku mõju ja juhuslik keskkonna mõju,

\bar{P} on keskmine fenotüübiväärtus vaatlusaluses populatsioonis,

A_{S_j} ja $A_{D_{jk}}$ on vastavalt j . isa aretusväärtus ja j . isa k . järglase ema aretusväärtus.

Et huvi pakub vaid isalt järglasel pärandunud geenide mõju, loetakse kõik ülejäänud geneetilised efektid kuuluvaks juhuslike keskkonnamõjude hulka: $\frac{1}{2}A_{D_{jk}} + MS_{jk} + E_{jk} = E_{jk}^*$.

Fenotüübiline dispersioon avaldub vastavalt lineaarsele geneetilisele mudelile (pt 3.1.2) kujul

$$\sigma_P^2 = \sigma_A^2 + \sigma_E^2 = \underbrace{\frac{1}{4}\sigma_{A_S}^2 + \frac{1}{4}\sigma_{A_D}^2 + \sigma_{MS}^2}_{\sigma_A^2} + \sigma_E^2 = \frac{1}{4}\sigma_{A_S}^2 + \sigma_E^2 . \quad (7.2)$$

Isa mudelit rakendades eeldatakse, et isade aretusväärtuste dispersioon on sama suur, kui kogu populatsiooni loomade aretusväärtuste dispersioon: $\sigma_{A_S}^2 \approx \sigma_A^2$.

Viimasest tõdemusest ja valemist (7.2) lähtuvalt on kogu aditiivgeneetiline dispersioon $\sigma_A^2 \approx 4 \times \sigma_{A_S}^2$ ja ainukese populatsioonigeneetilise parameetrina isa mudelist leitav päritavuskoeffitsient avaldub suhtena $h^2 = 4\sigma_{A_S}^2 / \sigma_P^2$ (vt valem (3.9)).

- **Statistilise mudelina** avaldub isa mudel kujul

$$y_{ijk} = \mu + H_i + s_j + e_{ijk}, \quad (7.3)$$

kus

μ on populatsiooni keskmine,

H_i on fikseeritud faktori taseme i mõju (nn karja-aasta-sesooni efekt), $i = 1, \dots, p$ (p on fikseeritud faktori tasemete arv),

s_j on isa j juhuslik mõju (geneetiline efekt), $j = 1, \dots, m$ (m on isade arv),

e_{ijk} on juhuslik viga, $k = 1, \dots, N$ ($N = \sum_{j=1}^m n_j$ on vaatluste arv, n_j on j . isa järglaste arv; kui järglaste arv kõigil isadel on n , siis avaldub vaatluste arv kujul $N = n \times m$).

Isa mõju s_j eeldatakse olevat normaaljaotusega keskvärtusega 0 ja dispersiooniga σ_s^2 ning juhuslik viga e_{ijk} normaaljaotusega keskvärtusega 0 ja dispersiooniga σ_e^2 . Seega, eeldatakse keskkonna ja genotüübi sõltumatust, avaldub kogu uuritava tunnuse dispersioon summana

$$\sigma_y^2 = \sigma_s^2 + \sigma_e^2. \quad (7.4)$$

Kõrvutades isa mudeli geneetilise esituse (7.1) statistilise esitusega (7.3) ning dispersiooni esituse (7.2) esitusega (7.4), on ilmne, et

- keskmine fenotüübiväärtus vaatlusaluses populatsioonis \bar{P} on statistilisest mudelist hinnatav suurusega $\mu + H_i$,
- statistilisest mudelist hinnatav isa mõju s_j kujutab enesest hinnangut poolele isa j aretusväärtusest A_{s_j} ,
- päritavuskoefitsient avaldub statistilise mudeli parameetrite kaudu kujul $h^2 = 4\sigma_s^2 / (\sigma_s^2 + \sigma_e^2)$.

- **Maatrikskujul** esitatakse isa mudel (7.3) järgmiselt:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s} + \mathbf{e} \quad (7.5)$$

kus

\mathbf{y} on uuritava tunnuse juhuslik $N \times 1$ vektor,

\mathbf{b} on fikseeritud efektide $p \times 1$ vektor,

\mathbf{s} on juhuslike isa mõjude $m \times 1$ vektor,

\mathbf{e} on juhuslike vigade $N \times 1$ vektor,

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

\mathbf{Z} on juhuslikele efektidele vastav $N \times m$ plaanimaatriks.

Keskvärtused ja dispersioonimaatriksid defineeritakse analoogselt peatükis 2.4 kirjeldatule:

$$E(\mathbf{s}) = 0, \quad E(\mathbf{e}) = 0, \quad E(\mathbf{y}) = \mathbf{X}\mathbf{b},$$

$$\mathbf{G} = \text{var}(\mathbf{s}) = \text{var} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} = \begin{pmatrix} \sigma_s^2 & 0 & \cdots & 0 \\ 0 & \sigma_s^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_s^2 \end{pmatrix}_{m \times m} = \mathbf{I}_m \sigma_s^2,$$

$$\mathbf{R} = \text{var}(\mathbf{e}) = \mathbf{I}_N \sigma_e^2,$$

$$\text{cov}(\mathbf{s}, \mathbf{e}^T) = 0$$

ning

$$\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{Z}^T \sigma_s^2 + \mathbf{I}_N \sigma_e^2.$$

Fakt, et $\text{var}(\mathbf{e}) = \mathbf{I}_N \sigma_e^2$, märgib vaatluste sõltumatust (see võib muutuda, kui ühele loomale vastab mitu mõõtmist) ning kõigi vaatluste eeldatavalt võrdset keskkonnatingimustest tingitud varieeruvust.

Keskkonna ja genotüübi sõltumatus on siin kirjas kujul $\text{cov}(\mathbf{s}, \mathbf{e}^T) = 0$.

Et $\text{var}(\mathbf{s}) = \mathbf{I}_m \sigma_s^2$, siis on kõigi geneetiliste efektide dispersioon σ_s^2 ja isade vaheline kovariatsioon on null (isad pole omavahel sugulased).

7.1.2 Aretusväärtuste hindamine isa mudelist

Vastavalt isa mudeli definitsioonile on selle alusel võimalik hinnata vaid isa mõju. Eeldusel, et järglaste fenotüübiväärtuste erinevust põhjustavad mittegeneetilised mõjud on arvesse võetud ja valikut emade osas ei ole toimunud, on isa mudelist hinnatava isa mõju näol tegu isalt järglasele pärandunud geenide summaarse mõjuga, mis moodustab poole isa aretusväärtusest. Sageli ei esitatagi poolõvede

modeli tulemusena mitte looma kogu aretusväärtust, vaid üksnes selle mudelist hinnatavat osa, so **järglasele päranduvat geneetilist efekti** (*transmission ability*).

Maatrikskujul esitatud isa mudelile (7.5) vastavad BLUP-hinnangud isa mõjule, arvestades karja-aasta-sesooni efekti, on leitavad Hendersoni segamudeli võrrandist

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + k\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}, \quad (7.6)$$

kus

$$k = \frac{\sigma_e^2}{\sigma_s^2} = \frac{4-h^2}{h^2}, \quad (7.7)$$

$$\text{sest } \frac{1}{h^2} = \frac{(\sigma_s^2 + \sigma_e^2)}{4\sigma_s^2} = \frac{1}{4} + \frac{\sigma_e^2}{4\sigma_s^2}, \text{ millest } \frac{\sigma_e^2}{\sigma_s^2} = 4 \times \left(\frac{1}{h^2} - \frac{1}{4} \right) = \frac{4-h^2}{h^2}.$$

Isa j aretusväärtuse a_j hinnang on leitav valemist (7.6) saadud isa mõju hinnangu \hat{s}_j kaudu:

$$\hat{a}_j = 2 \times \hat{s}_j.$$

Juhul, kui päritavuskoeffitsiendi väärtus h^2 pole teada, asendatakse selle väärtus valemis (7.6) hinnanguga \hat{h}^2 , mis on leitud statistilisest mudelist hinnatud dispersioonikomponentide kaudu:

$$\hat{h}^2 = 4\hat{\sigma}_s^2 / (\hat{\sigma}_s^2 + \hat{\sigma}_e^2).$$

Näide 7.1. Lehmade piimatoodangut on mõõdetud kahes erinevas karjas. Kõik mõõtmised on tehtud samal aastal ja kõik vaadeldud lehmad on ühevanused. Lisaks karjale, kuhu lehm kuulub, on teada ka tema isa. Ülesandeks on hinnata isade aretusväärtust, kasutades isa mudelit. Piimatoodangu päritavuseks võib lugeda 0,25. Andmetabel on järgmine:

Lehm	Isa	Kari	Piimatoodang
101	1	1	7000
102	2	1	6600
103	3	1	6800
104	3	1	7200
201	2	2	8000
202	2	2	8300
203	3	2	8900

Mudel sisaldab karja mõju kui fikseeritud efekti ning isa mõju kui juhuslikku efekti:

$$\mathbf{b} = \begin{pmatrix} kari_1 \\ kari_2 \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} isa_1 \\ isa_2 \\ isa_3 \end{pmatrix}.$$

Lahendite ühesuse huvides loeme vabaliikme μ võrdseks nulliga. Plaanimaatriksid \mathbf{X} ja \mathbf{Z} , mis seovad iga vaatluse talle mõjuvate faktorite tasemetega, on

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Faktorite mõjude hindamiseks paneme esmalt kirja mudelile vastavad vähimruutude võrrandid (6.12) ning liidame seejärel segamudeli võrrandite saamiseks isamõjule vastavaile diagonaalelementidele dispersioonide suhte k . Meie andmetele vastavad vähimruutude võrrandi kordajate maatriks ja vabaliikmete vektor on

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} \end{pmatrix} = \begin{pmatrix} 4 & 0 & 1 & 1 & 2 \\ 0 & 3 & 0 & 2 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 3 & 0 \\ 2 & 1 & 0 & 0 & 3 \end{pmatrix} \text{ ja } \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix} = \begin{pmatrix} 27600 \\ 25200 \\ 7000 \\ 22900 \\ 22900 \end{pmatrix}.$$

Päritavuskoeffitsient $h^2 = 0,25$ ning seega $k = \frac{\sigma_e^2}{\sigma_s^2} = \frac{1 - \frac{1}{4}h^2}{\frac{1}{4}h^2} = 15$. Faktorite mõjude hinnangud on nüüd leitavad segamudeli võrrandist

$$\begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \end{pmatrix} = \begin{pmatrix} 4 & 0 & 1 & 1 & 2 \\ 0 & 3 & 0 & 2 & 1 \\ 1 & 0 & 1+15 & 0 & 0 \\ 1 & 2 & 0 & 3+15 & 0 \\ 2 & 1 & 0 & 0 & 3+15 \end{pmatrix}^{-1} \begin{pmatrix} 27600 \\ 25200 \\ 7000 \\ 22900 \\ 22900 \end{pmatrix} = \begin{pmatrix} 6890 \\ 8418 \\ 6,9 \\ -45,9 \\ 39,0 \end{pmatrix}.$$

Esimese isa aretusväärtuse hinnang on $\hat{a}_1 = 2 \times \hat{s}_1 = 2 \times 6,9 = 13,8$; teisel isal $\hat{a}_2 = -91,8$ ja kolmandal isal $\hat{a}_3 = 78$.

Märkus. Et vastavalt juhuslike efektide olemusele võrdub nende keskvärtus nulliga, siis võrduvad nulliga ka isade keskmine mõju ja isade aretusväärtuste summa. Meie näite korral $\hat{a}_1 + \hat{a}_2 + \hat{a}_3 = 13,8 + (-91,8) + 78 = 0$.

Küsimus. Nii teise kui ka kolmanda isa 3 tütre summaarne toodang on 22900 kg. Miks on aga nende hinnatud aretusväärtused nii erinevad?

Kui kõigi analüüsitava isade järglased on pärit sarnastest keskkonnatingimustest või on fenotüübiväärtused analüüsi eelselt mittegeneetiliste mõjude suhtes korrigeeritud, lihtsustub mudel (7.3) kujule

$$y_{jk} = \mu + s_j + e_{jk}.$$

Isa j mõju hinnang avaldub taolise mudeli puhul võrrandist (6.26) ja aretusväärtus siinses peatükis kasutatud tähistuste abil kujul

$$\hat{a}_j = 2 \times \hat{s}_j = \frac{2 \times n_j \sigma_s^2}{\sigma_e^2 + n_j \sigma_s^2} (\bar{y}_j - \hat{\mu}), \quad (7.8)$$

kus \bar{y}_j on j . isa järglaste keskmine fenotüübiväärtus ja n_j järglaste arv. Et suurus $(\bar{y}_j - \hat{\mu})$ kujutab enesest j . isa järglaste keskmist erinevust populatsiooni/karja keskmisest ning selle erinevuse kordaja on tänu seosele (7.7) teisenatav kujule

$$\frac{2n_j \sigma_s^2}{\sigma_e^2 + n_j \sigma_s^2} = \frac{2n_j}{(\sigma_e^2 + n_j \sigma_s^2) / \sigma_s^2} = \frac{2n_j}{\sigma_e^2 / \sigma_s^2 + n_j} = \frac{2n_j}{n_j + (4 - h^2) / h^2},$$

avalduv j . isa aretusväärtus tema järglaste keskmise fenotüübiväärtuse kaudu valemiga

$$\hat{a}_j = \frac{2n_j}{n_j + (4 - h^2) / h^2} (\bar{y}_j - \hat{\mu}).$$

Saadud tulemus on identne peatükis 5.2.3 tuletatud selektsiooniindeksiga (valemid (5.16) ja (5.18)) hindamaks isa aretusväärtust tema järglastel mõõdetud fenotüübiväärtuste alusel. Seega on tehtud eeldustel (keskkonnamõjude puudumine) selektsiooniindeksi kujul leitud aretusväärtuste parimad lineaarsed prognoosid (BLP) identsed üldisest lineaarsest mudelist leitud aretusväärtuste parimate lineaarsete nihketa prognoosidega (BLUP).

7.1.3 Isa mudelist leitud geneetiliste parameetrite hinnangute täpsusest

Isa mudeli lihtsus, mis on peamine põhjus, miks seda mudelit lisaks isade aretusväärtuste hindamisele küllaltki laialdaselt ka uuritava tunnuse geneetilist determineeritust terves populatsioonis kirjeldava päritavuskoeffitsiendi hindamiseks kasutatakse (eriti juhul, kui tegu on uudse näitajaga ja/või populatsiooniga/tõuga), annab võimaluse uurida ka erinevate katseplaanide/andmestruktuuride mõju geneetiliste parameetrite hinnangute täpsusele.

Statistilisest mudelist leitud mingi parameetri hinnangu täpsust näitab see, kui palju või vähe leitud hinnang varieerub. Viimast kirjeldatakse enamasti hinnangu standardhälbega (või hinnangu standardhälbe hinnanguga, mis on defineeritud kui **standardviga**), näiteks päritavuskoeffitsiendi hinnangu täpsust mõõdab suurus $\sigma(\hat{h}^2)$ (või $se(\hat{h}^2) = \sigma(\hat{h}^2)$). Mudelist leitud prognooside täpsust mõõdab **keskmine ruutviga** (*mean square error*, MSE) või ruutjuur viimasest, mille hinnangut sageli (näiteks statistikaprogrammide väljatrukkides) ka lihtsalt standardveaks nimetatakse. Näiteks isa mudeli abil saadud isa j mõju hinnangu täpsust mõõdab suurus $\sqrt{MSE(\hat{s}_j)} = \sigma(\hat{s}_j - s_j)$.

Hoolimata isa mudeli lihtsusest, on selle abil leitud parameetrite hinnangute standardvigade matemaatilised avaldised suhteliselt keerukad. Siiski on arvutite abil võimalik läbi viia simulatsiooni-

eksperimente, uurimaks erinevate andmestruktuuride mõju hinnangute täpsusele, ja leida ka optimaalne tütarde arv isa kohta ette antud vaatluste arvu ja päritavuskoeffitsiendi väärtuse korral.

Olles fikseerinud tulevase andmestiku suuruse N ja oletatava päritavuskoeffitsiendi väärtuse h^2 , on optimaalne tütarde arv isa kohta isade aretusväärtuste hindamise soovi korral esitatav valemiga

$$n = \frac{-1 + \sqrt{1 + N\tau}}{\tau}, \quad (7.9)$$

kus $\tau = \sigma_s^2 / \sigma_e^2 = h^2 / (4 - h^2)$. Soovides aga isa mudelist hinnata võimalikult täpselt päritavuskoeffitsienti, tuleks andmestik moodustada, võttes iga isa järglaste arvaks

$$n = \frac{N(\tau + 1) + 1}{N\tau + 2}. \quad (7.10)$$

Näide 7.2. Soovitakse uurida lehmade söödakasutuse efektiivsuse geneetilist determineeritust. Raha on eraldatud nii palju, et sellest piisab 400 looma pidevaks jälgimiseks, mõõtmiseks ja katsete läbi viimiseks. Kui mitmelt isalt ja kui palju peaks võtma katsesse tütreid, et hinnata päritavuskoeffitsienti võimalikult täpselt?

Planeeritav andmestiku suurus $N = 400$. Teades teiste maade teadlaste uuringuist, et analüüsitava parameetri päritavuskoeffitsiendi väärtus on ligikaudu 0,4, saame optimaalseks tütarde arvaks isa kohta vastavalt valemile (7.10)

$$n = \frac{N(\tau + 1) + 1}{N\tau + 2} = \frac{N \left[\frac{h^2}{(4-h^2)} + 1 \right] + 1}{N \frac{h^2}{(4-h^2)} + 2} = \frac{400 \left[\frac{0,4}{(4-0,4)} + 1 \right] + 1}{400 \frac{0,4}{(4-0,4)} + 2} = 9,59 \approx 10.$$

Seega oleks uuritava tunnuse päritavuskoeffitsiendi hindamiseks vaatlusaluses karjas/populatsioonis antud ajahetkel mõistlik valida välja $N/n = 400/10 = 40$ pulli igaüks 10 tütrega.

Kui nüüd aga lugeda päritavuskoeffitsiendi väärtus teadaolevaks ja püstitada küsimus nii: kui mitu tütar peaks igalt pullilt analüüsima, saamaks võimalikult täpseid hinnanguid pullide aretusväärtustele? Vastuse saame leida valemist (7.9):

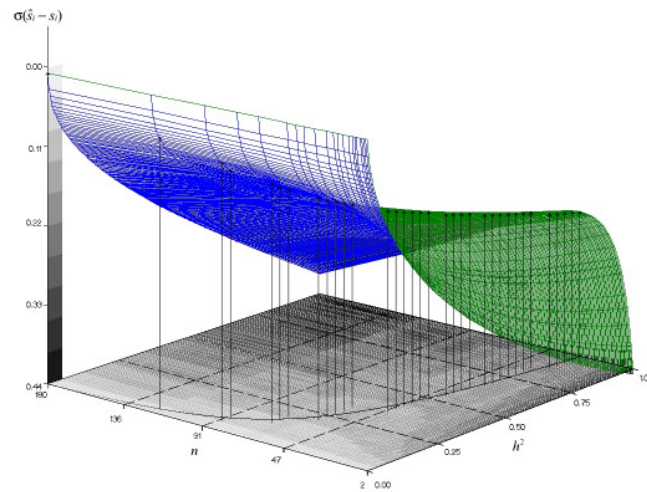
$$n = \frac{-1 + \sqrt{1 + N\tau}}{\tau} = \frac{-1 + \sqrt{1 + N \frac{h^2}{(4-h^2)}}}{\frac{h^2}{(4-h^2)}} = \frac{-1 + \sqrt{1 + 400 \frac{0,4}{(4-0,4)}}}{\frac{0,4}{(4-0,4)}} = 51,67 \approx 52.$$

Et ette antud loomade arv 400 52-ga täpselt ei jagu, siis tuleks võtta uuringusse 8 pulli igaüks 50 tütrega.

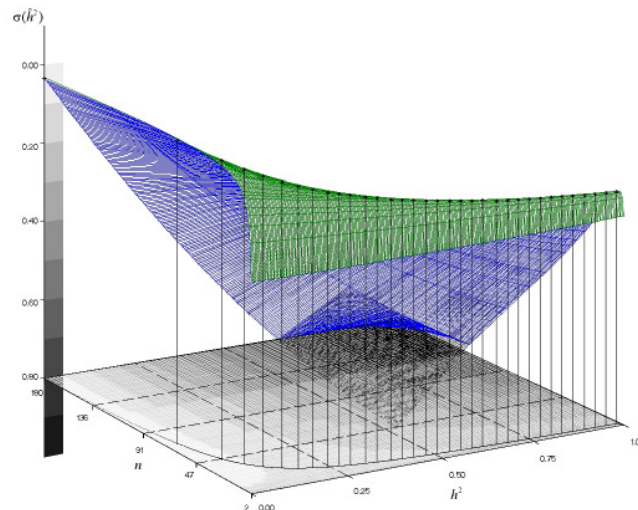
Joonistel 7.1 ja 7.2 on kujutatud vastavalt isa mõju ja päritavuskoeffitsiendi hinnangute täpsus (mõõdetuna hinnangu standardhälvena) sõltuvalt päritavuskoeffitsiendi väärtusest ja tütarde arvust isa kohta fikseeritud andmestiku suuruse $N = 360$ korral.

- Osutub, et mida väiksem on uuritava tunnuse aditiivgeneetiline determineeritus (st, et mida väiksem on päritavuskoeffitsiendi väärtus), seda suuremat tütarde arvu isa kohta on vaja.
- Päritavuskoeffitsiendi hindamiseks on vaja pigem suurt hulka isasid väiksema arvu tütardega.
- Aretusväärtuste hindamiseks on vaja jällegi suuremat hulka tütreid.

Viimased seosed on muidugi loomulikud, sest päritavuskoeffitsiendi väärtus sõltubki eelkõige isade erinevusest – mida enam isasid, seda paremini on võimalik erinevus tuvastatav, samas kui isa aretusväärtus sõltub tema tütarde keskmisest fenotüübiväärtusest – mida enam tütreid, seda täpsemalt on nende keskmine fenotüübiväärtus hinnatav.



Joonis 7.1. Isa mõju hinnangu täpsus (mõõdetuna keskmise prognoosiveana $\sigma(\hat{s}_i - s_i)$) sõltuvalt järglaste arvust ja päritavuskoefitsiendi väärtusest vaatluste arvu $N = 360$ korral. Pidev joon xy -tasandil märgib optimaalset tütarde arvu isa kohta pideval ja vertikaalsed jooned diskreetsel skaalal.



Joonis 7.2. Päritavuskoefitsiendi hinnangu täpsus (mõõdetuna hinnangu standardhällbena $\sigma(\hat{h}^2)$) sõltuvalt järglaste arvust ja tegelikust päritavuskoefitsiendi väärtusest vaatluste arvu $N = 360$ korral. Pidev joon xy -tasandil märgib optimaalset tütarde arvu isa kohta pideval ja vertikaalsed jooned diskreetsel skaalal.

7.1.4 Täisõvede mudel

Juhul, kui indiviidi aretusväärtuse hinnang baseerub täisõvede analüüsil (analüüsitakse mõlema vanema aditiivset geneetilist mõju järglastele), on kasutatav mudel eelnevast pisut erinev. Nimelt peab arvestama, et samasse täisõvede gruppi kuuluvad loomad sarnanevad teineteisele vähemalt kolmel põhjusel. Need on ühine isa, ühine ema ning samad sünnieelsed ja vahel ka sünnijärgsed kasvutingimused. Viimased võetakse geneetilisest mudelist kokku ühte, nn **ühist keskkonna mõju** iseloomustavasse liidetavasse E_C .

- Eelnevat arvesse võttes avaldub isa i ja ema j järglase k fenotüübiväärtus P_{ijk} **geneetilise mudelina** järgmiselt:

$$P_{ijk} = \bar{P} + \frac{1}{2} A_{S_i} + \frac{1}{2} A_{D_j} + MS_{ijk} + E_{C_{ij}} + E_{ijk} . \quad (7.11)$$

Siin

- \bar{P} on keskmine fenotüübiväärtus vaatlusaluses populatsioonis,
- A_{S_i} ja A_{D_j} on vastavalt i . isa aretusväärtus ja i . isaga ristatud j . ema aretusväärtus,
- $E_{C_{ij}}$ on i . isa ja j . ema kõigile järglastele mõjunud ühised keskkonna mõjud,
- MS_{ijk} ja E_{ijk} on vastavalt Mendeli valiku mõju ja juhuslik keskkonna mõju.

Seejuures on Mendeli valiku mõju täisõvedele erinev (kui see nii ei ole, on ilmselt tegu ühemuna-kaksikutega).

▪ **Statistilise mudeli** kirjapanekul arvestatakse enamasti, et üks isaloom on paaritatud mitme emasega ning tihti on andmete struktuur hierarhiline, st iga ema on paaritatud vaid ühe kindla isaga. Sellisel juhul on ema j mõju $d_{j(i)}$ allutatud isa i mõjule s_i . Täisõvede analüüsi statistiline mudel on järgmine:

$$y_{ijk} = \mu + s_i + d_{j(i)} + e_{ijk}. \quad (7.12)$$

Kõrvutades täisõvede mudeli geneetilise esituse (7.11) statistilise esitusega (7.12) ilmneb, et andmetest hinnatav ei ole ei Mendeli valiku mõju MS_{ijk} ega ka püsiv keskkonnamõju $E_{C_{ij}}$ – neist esimene ei ole statistilises mudelis eristatav mudeli jääkliikmest, teine aga ema mõjust. Seega on statistilisest mudelist võimalik hinnata vaid isa mõju.

▪ Fenotüübidispersioon σ_P^2 avaldub geneetilisest mudelist summana

$$\sigma_P^2 = \sigma_A^2 + \sigma_{E_c}^2 + \sigma_E^2 = \frac{1}{4}\sigma_{A_s}^2 + \frac{1}{4}\sigma_{A_d}^2 + \sigma_{MS}^2 + \sigma_{E_c}^2 + \sigma_E^2.$$

Sarnaselt isa mudelile eeldatakse, et isade aretusväärtuste dispersioon on sama suur, kui kogu populatsiooni loomade aretusväärtuste dispersioon: $\sigma_{A_s}^2 \approx \sigma_A^2$, sama eeldus tehakse ka emade aretusväärtuste dispersiooni kohta: $\sigma_{A_d}^2 \approx \sigma_A^2$. Seega on ka poolõvede mudeli korral pool aditiivgeneetilisest varieeruvusest tingitud Mendeli valikust ja ei ole statistilise mudeliga hinnatav. Et püsiv keskkonnamõju ei ole eraldi vaadeldav ema mõjust, ei saa eraldi hinnata ka püsivast keskkonna mõjust tingitud varieeruvust. Seega võib fenotüübidispersiooni komponendid grupeerida:

$$\sigma_P^2 = \frac{1}{4}\frac{\sigma_{A_s}^2}{\sigma_A^2} + \left[\frac{1}{4}\frac{\sigma_{A_d}^2}{\sigma_A^2} + \sigma_{E_c}^2 \right] + \left[\frac{\sigma_{MS}^2 + \sigma_E^2}{\frac{1}{2}\sigma_A^2} \right]. \quad (7.13)$$

Statistilisest mudelist avaldub uuritava tunnuse dispersioon summana

$$\sigma_y^2 = \sigma_s^2 + \sigma_d^2 + \sigma_e^2, \quad (7.14)$$

kus isa mõjude dispersioon mõõdab $\frac{1}{4}$ kogu aditiivgeneetilisest dispersioonist, ema mõjude dispersioon sisaldab nii $\frac{1}{4}$ aditiivgeneetilisest dispersioonist kui ka ühisest keskkonnamõjust tingitud varieeruvust.

▪ Kõrvutades esitusi (7.13) ja (7.14) võib välja tuua 3 statistilisest mudelist hinnatavat populatsiooni geneetilist ja mittegeneetilist determineeritud kirjeldavat parameetrit:

– päritavus, mis mõõdab aditiivgeneetilise varieeruvuse osa kogu dispersioonist, on statistilisest mudelist arvatav valemiga

$$h^2 = 4\sigma_s^2 / \sigma_y^2$$

(päritavus arvutatakse sarnaselt isamudelile vaid isast tingitud varieeruvuse alusel, emale vastava dispersiooni arvestamine toob kaasa päritavuskoeffitsiendi nihkega hinnangu – mõlemas veel kõne alla tulevas h^2 arvutamise valemis $h^2 = 2(\sigma_s^2 + \sigma_d^2) / \sigma_y^2$ ja $h^2 = 4\sigma_d^2 / \sigma_y^2$ sisaldab lugeja peale geneetilist varieeruvust kajastavate liikmete σ_s^2 ja σ_d^2 ka püsivast keskkonna mõjust tingitud dispersiooni, sest vastavalt valemile (7.13) ja (7.14) $\sigma_d^2 \approx \frac{1}{4}\sigma_A^2 + \sigma_{E_c}^2$);

– ühisest keskkonna mõjust tingitud varieeruvuse osakaal koguarveeruvuses $\sigma_{E_c}^2 / \sigma_P^2$ on statistilisest mudelist hinnatav suhtena

$$c^2 = (\sigma_d^2 - \sigma_s^2) / \sigma_y^2;$$

– juhuslikust keskkonna mõjust tingitud varieeruvuse osakaal koguarveeruvuses σ_E^2 / σ_P^2 on statistilisest mudelist hinnatav suhtena

$$e^2 = (\sigma_e^2 - 2\sigma_s^2) / \sigma_y^2.$$

Eeltoodud definitsioonide põhjal on selge, et $h^2 + c^2 + e^2 = 1$.

7.2 LOOMA MUDEL

Looma mudel (*animal model*) sisaldab geneetilise efektna looma enda aretusväärtust ning tema lihtsaim esitusviis on

$$y_{ij} = \mu + H_i + a_j + e_{ij}, \quad (7.15)$$

kus

μ on üldkeskmine,

H_i on nn. karja-aasta-sesooni efekti i mõju,

a_j on looma j aretusväärtus,

e_{ij} on karjast i pärit loomale j vastav juhuslik keskkonna mõju.

Lisades mudelisse informatsiooni loomade suguluse kohta, on seeläbi võimalik hinnata aretusväärtused kõigile andmestikku kuuluvatele loomadele.

Lisaks võimaldab kõikvõimalike sugulussidemete kompleksne arvestamine mõõta inbriidingu ja selektsiooni mõju ning seeläbi saada objektiivsemad hinnangud geneetilistele parameetritele.

7.2.1 Aditiivgeneetilise suguluse maatriks

Fakt, et kaks looma on sugulased, ilmneb selles, et nendevaheline geneetilisest faktoritest tingitud kovariatsioon on nullist erinev. Oluliseim osa loomadevahelisest geneetilisest varieeruvusest on põhjustatud geenide aditiivse toime variatsioonist, ülejäänud, küllalt väheoluline osa, on seletatav dominantsusest (alleelide interaktsioon) ja epistaasist (erinevetes lookustes asuvate geenide vastastikune mõju) tingitud dispersiooniga.

Aditiivgeneetiline sugulus loomade vahel on väljendatav **aditiivgeneetilise suguluse maatriksi** (*additive genetic relationship matrix*) **A** abil.

Maatriks **A** on $n \times n$ maatriks, kus n on uuritavate loomade arv. Maatriksi **A** elementideks a_{ij} on peatükis 5 käsitletud aditiivgeneetilise suguluse kordajad, mis on võrdsed kahekordse tõenäosusega, et juhuslik alleel indiviidi i genotüübist on päritolult identne vastava alleeliga indiviidi j genotüübist.

Arvutuslikult lihtsaim on maatriksi **A** konstrueerimine rekursiivsel teel kasutades Hendersoni meetodit (Pt 4.4). Selleks järjestatakse individid geneoloogiliselt, st vanemad peavad paiknema eespool järglasi. Tähistagu nüüd s_i indiviidi i isa ja d_i indiviidi i ema. Siis on maatriksi **A** elemendid leitavad järgmiste reeglite alusel.

- Kui looma i mõlemad vanemad on teadmata, siis

$$a_{ii} = 1 \text{ ja } a_{ij} = a_{ji} = 0, \quad j = 1, \dots, i-1.$$

- Kui vaid looma i üks vanemaist (isa või ema) on teada, siis

$$a_{ii} = 1 \text{ ja } a_{ij} = a_{ji} = \frac{1}{2}(a_{is_j}) \text{ või } a_{ij} = a_{ji} = \frac{1}{2}(a_{id_j}), \quad j = 1, \dots, i-1$$

(a_{id_j} või a_{is_j} võrdub nulliga – see tuleneb eeldusest, et teadmata vanemad loetakse suguluses mitteolevaiks).

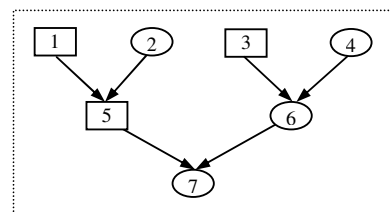
- Kui looma i mõlemad vanemad (isa ja ema) on teada, siis

$$a_{ii} = 1 + \frac{1}{2}a_{s_i d_i} \text{ ja } a_{ij} = a_{ji} = \frac{1}{2}(a_{is_j} + a_{id_j}), \quad j = 1, \dots, i-1.$$

Maatriksi **A** konstrueerimist alustatakse ülemisest vasakust nurgast (vanimast loomast), liikudes seejärel ühe rea ja veeru kaupa mööda diagonaali allapoole ning täites igal sammul maatriksi read ja veerud kuni diagonaalini (leides konkreetse looma suguluse endast vanemate loomadega).

Näide 7.3. Olgu meil järgmised skemaatiliselt ja tabelikujul esitatud põlvnemisandmed:

Loom	Isa	Em
4	1	2
5	1	3
6	4	5



Indiviidid järjestatakse nii, et vanemad paiknevad eespool järglasi.

Loom	Isa	Ema
1	-	-
2	-	-
3	-	-
4	1	2
5	1	3
6	4	5

Kuna loomade 1, 2 ja 3 vanemad on teadmata, kirjutatakse maatriksi \mathbf{A} diagonaalile vastavasse kohta üks (looma aditiivgeneetiline sugulus iseendaga on üks) ja täidetakse vastav rida ja veerg kuni diagonaalini nullidega.

$$\mathbf{A} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ \hline & & & & & \\ & & & & & \\ & & & & & \end{array} \right) \end{matrix}$$

Liikudes ridahaaval põlvnemistabelis allapoole, saame täidetud kogu maatriksi \mathbf{A} :

$$\mathbf{A} = \begin{matrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0,5 & 0,5 & 0,5 \\ 0 & 1 & 0 & 0,5 & 0 & 0,25 \\ 0 & 0 & 1 & 0 & 0,5 & 0,25 \\ \hline 0,5 & 0,5 & 0 & 1 & 0,25 & 0,625 \\ 0,5 & 0 & 0,5 & 0,25 & 1 & 0,625 \\ 0,5 & 0,25 & 0,25 & 0,625 & 0,625 & 1,125 \end{array} \right) \end{matrix}$$

Näiteks,

$$a_{41} = a_{14} = \frac{1}{2}(a_{11} + a_{12}) = \frac{1}{2}(1 + 0) = 0,5;$$

$$a_{42} = a_{24} = \frac{1}{2}(a_{21} + a_{22}) = \frac{1}{2}(0 + 1) = 0,5;$$

$$a_{43} = a_{34} = \frac{1}{2}(a_{31} + a_{32}) = \frac{1}{2}(0 + 0) = 0; \quad a_{44} = 1 + \frac{1}{2}a_{45} = 1 + \frac{1}{2} \times 0 = 1;$$

$$a_{65} = a_{56} = \frac{1}{2}(a_{54} + a_{55}) = \frac{1}{2}(0,25 + 1) = 0,625; \quad a_{66} = 1 + \frac{1}{2} \times a_{45} = 1 + \frac{1}{2} \times 0,25 = 1,125.$$

7.2.2 Looma mudel maatrikskujul

Looma mudel on maatrikskujul kirjapandav analoogselt isa mudelile (7.5):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \quad (7.16)$$

Siin \mathbf{a} on juhuslike aretusväärtuste $n \times 1$ vektor (n on kõigi andmestikku kuuluvate loomade arv), ülejäänud muutujad on defineeritud nagu isa mudeli korral.

Keskväärtused ja dispersioonimaatriksid on:

$$\begin{aligned} E(\mathbf{a}) &= 0, \quad E(\mathbf{e}) = 0, \quad E(\mathbf{y}) = \mathbf{Xb}, \\ \text{var}(\mathbf{a}) &= \mathbf{G} = \mathbf{A}\sigma_a^2, \\ \text{var}(\mathbf{e}) &= \mathbf{R} = \mathbf{I}_N\sigma_e^2, \\ \text{cov}(\mathbf{a}, \mathbf{e}^T) &= 0 \end{aligned}$$

ning

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZAZ}^T\sigma_a^2 + \mathbf{I}\sigma_e^2$$

(σ_a^2 on aditiivgeneetiline dispersioon). Et $\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2$, siis erinevalt isa mudelist on arvestatud ka loomadevahelisi sugulussidemeid.

Segamudeli võrrand saab kuju

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + k\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}, \quad (7.17)$$

kus $k = \sigma_e^2 / \sigma_a^2 = (1 - h^2) / h^2$.

NB! Kui vektoris \mathbf{a} esindatud loomad pole omavahel sugulased, siis $\mathbf{A} = \mathbf{A}^{-1} = \mathbf{I}$.

Näide 7.4. Olgu meil järgmine andmetabel:

Lehm	Em	Isa	Uuritav tunnus
2	-	-	11
4	-	-	7
5	2	1	10
6	2	1	8
7	4	3	9

Lehmade 2 ja 4 vanemad on teadmata. Lehmad 5 ja 6 on ET-tütred (täisõed). Lehma 2 eelistati järglaspõlvkonna tootmisel kuna temal mõõdetud tunnuse väärtus oli parem. Pullidel uuritavat tunnust mõõdetud pole. Uuritava tunnuse päritavus (h^2) on 0,5. Lihtsuse mõttes on eeldatud, et ainus mudelisse lülitatud fikseeritud efekt on üldkeskmise. Soovime kasutada looma mudelit, valimaks välja geneetiliselt potentsiaalilt parimaid isendeid populatsiooni taastootmiseks.

Hindame aretusväärtused kahel viisil: esmalt ignoreerime loomadevahelisi sugulusidemeid ja seejärel võtame need arvesse, ning vaatame, mil määral nende arvestamine tulemusi muudab.

Mudel indekskujul kirjapanduna on järgmine:

$$y_i = \mu + a_i + e_i,$$

kus μ on üldkeskmise, a_i on looma i aretusväärtus, e_i on loomale i vastav juhuslik kesk-konna mõju.

Plaanimaatriksid on

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{ja} \quad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

ning vähimruutude võrrand esitub kujul

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ a_2 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} 45 \\ 11 \\ 7 \\ 10 \\ 8 \\ 9 \end{pmatrix}.$$

$$\text{Et } h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} = 0,5, \text{ siis } k = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2} = 1.$$

(1) Ignoreerides loomadevahelist sugulust ($\mathbf{A} = \mathbf{I}$), on segamudeli võrrand järgmine:

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1+1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1+1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1+1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1+1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1+1 \end{pmatrix} \begin{pmatrix} \mu \\ a_2 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} 45 \\ 11 \\ 7 \\ 10 \\ 8 \\ 9 \end{pmatrix}$$

ning selle lahendamisel saadud hinnangute vektor on

$$\begin{pmatrix} \hat{\mu} \\ \hat{a}_2 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \end{pmatrix} = \begin{pmatrix} 9 \\ 1 \\ -1 \\ 0,5 \\ -0,5 \\ 0 \end{pmatrix}.$$

Märkused.

- ✓ Kõigi indiviidide keskmine aretusväärtus võrdub nulliga.
- ✓ Leitud üldkeskmise $\mu = 9$ abil võime iga indiviidi aretusväärtuse arvutada valemist $\hat{a}_i = h^2(y_i - 9)$, kus y_i on uuritava tunnuse väärtus antud loomal.

(2) Võtame nüüd arvesse ka loomadevahelise suguluse, mis võimaldab arvestada selektsiooni ja inbriidingu võimalikku mõju ning hinnata aretusväärtused ka ilma vaatlusteta isadele. Selleks konstrueerime aditiivse geneetilise suguluse maatriksi \mathbf{A} kõigi andmestiku kuuluvate loomade (ka isade) jaoks.

$$\mathbf{A} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 \end{pmatrix}.$$

Segamudeli võrrandi (7.17) kirjapanekuks vajalik maatriksi \mathbf{A} pöördmaatriksi \mathbf{A}^{-1} on

$$\mathbf{A}^{-1} = \begin{pmatrix} 2 & 1 & 0 & 0 & -1 & -1 & 0 \\ 1 & 2 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1\frac{1}{2} & \frac{1}{2} & 0 & 0 & -1 \\ 0 & 0 & \frac{1}{2} & 1\frac{1}{2} & 0 & 0 & -1 \\ -1 & -1 & 0 & 0 & 2 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 2 \end{pmatrix}.$$

Kasutades suguluse maatriksit, hinnatakse aretusväärtused ka isadele (a_1 ja a_3). Need efektid on aga varem leitud plaanimaatriksis \mathbf{Z} arvestamata. Teades, et plaanimaatriksis \mathbf{Z} on üks rida iga vaatluse jaoks (siin näites 5) ning üks veerg iga geneetilise efekti jaoks (siin näites 7, so. 5 lehma, kellel on tunnus mõõdetud ja 2 isa), esitub \mathbf{Z} kujul

$$\mathbf{Z} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Vähimruutude võrrand on nüüd

$$\begin{pmatrix} 5 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} 45 \\ 0 \\ 11 \\ 0 \\ 7 \\ 10 \\ 8 \\ 9 \end{pmatrix}.$$

Liites $k\mathbf{A}^{-1}$ vähimruutude võrrandi kordajate maatriksile vastavale alammaatriksile, saame segamudeli võrrandi, kus on arvestatud ka loomadevahelisi sugulussidemeid:

$$\begin{pmatrix} 5 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 0 & 0 & -1 & -1 & 0 \\ 1 & 1 & 3 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1\frac{1}{2} & \frac{1}{2} & 0 & 0 & -1 \\ 1 & 0 & 0 & \frac{1}{2} & 2\frac{1}{2} & 0 & 0 & -1 \\ 1 & -1 & -1 & 0 & 0 & 3 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} 45 \\ 0 \\ 11 \\ 0 \\ 7 \\ 10 \\ 8 \\ 9 \end{pmatrix}.$$

Selle võrrandi lahendid on

$$\begin{pmatrix} \hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \end{pmatrix} = \begin{pmatrix} 8,926 \\ -0,198 \\ 0,938 \\ 0,148 \\ -0,889 \\ 0,605 \\ -0,062 \\ -0,222 \end{pmatrix}.$$

Märkused.

✓ Baasloomade, s.o. loomade, kelle vanemad on teadmata, aretusväärtuste hinnangute keskmine võrdub nulliga: $\frac{1}{4}(\hat{a}_1 + \hat{a}_2 + \hat{a}_3 + \hat{a}_4) = 0$. See on tingitud eeldusest, et baasloomad on juhuslikult valitud baaspopulatsioonist.

✓ Järglaspõlvkonna aretusväärtuste hinnangute keskmine on suurem kui null: $\frac{1}{3}(\hat{a}_5 + \hat{a}_6 + \hat{a}_7) = 0,11$. See väljendab nii juhusliku geenitriivi kui ka selektsiooni mõju. Meie näites annab kõrgema fenotüübiväärtusega ema kaks ja madalama fenotüübiväärtusega ema ühe järglase. Sellise valiku tulemus peegeldub järglaste keskmises aretusväärtuses.

✓ Võrdlus isamudeliga. Isade aretusväärtuste leidmisel isamudeli abil ignoreeritakse emade andmeid ning isa 1 kaks järglast (lehmad 5 ja 6) loetakse seeläbi poolõdedeks. Segamudeli võrrand isade mõjude prognoosimiseks oleks:

$$\begin{pmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2+7 & 0 \\ 1 & 0 & 1+7 \end{pmatrix}^{-1} \begin{pmatrix} 27 \\ 18 \\ 9 \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \\ 0 \end{pmatrix},$$

millest $\hat{a}_1 = 0$ ja $\hat{a}_3 = 0$. Seega on isa 1 ja isa 3 hinnatud geneetiliselt potentsiaalilt võrdseteks, mis on aga ilmselt vale.

7.3 ÜLESANDED

Võtke aluseks ülesande lõpus toodud andmestik, andke see sobival kujulⁱ ette aretusväärtuste hindamiseks kasutatavale *Excel*'i programmile 'SAMod_1c.xls' ja vastake järgmistele küsimustele.

- Hinnake kõigi täkkude tõutüübi aretusväärtused isa mudelist, võttes arvesse ka hindaja võimaliku mõju. Päritavuskoefitsiendi h^2 väärtuseks võtke 0,25.
 - Milline on hindajate vaheline erinevus?
 - Kuidas muutus täkkude paremusjärjestus võrreldes selektsiooniindeksite abil hinnatud aretusväärtustegaⁱⁱ. Miks taoline erinevus ilmnes?
- Hinnake kõigi hobuste tõutüübi aretusväärtused looma mudelist, võttes täiendavalt arvesse ka seda, et Tiktora 697 E on Tukker 703 E isa ning Rosett 600 E ja Rommik 710 E on sama ema järglased.
 - Kas ja kuidas see täkkude aretusväärtuste hinnanguid muudab? Miks?
- Kuivõrd muutuvad isade aretusväärtuste hinnangud, kui looma mudelis võtta päritavuskoefitsiendi väärtuseks 0,6?

Hobuse nimi	Isa	Hindaja	Tõutüüp
Vaida	Vigur 682 E	1	7
Vaara	Vigur 682 E	1	7
Villu	Vigur 682 E	2	7
Viker	Vigur 682 E	1	6
Teini	Tukker 703 E	2	8
Taara	Tukker 703 E	1	7
Tenor	Tiktora 697 E	1	6
Tommy	Tiktora 697 E	1	7
Relli	Rosett 600 E	2	7
Rollu	Rosett 600 E	1	6
Raul	Rommik 710 E	2	8
Rolf	Rommik 710 E	2	9
Ruubik	Rommik 710 E	1	5

Esitada ei ole vaja üksikasjalikke arvutusi, piisab, kui on ära toodud erinevate mudelitega hinnatud täkkude aretusväärtused, nendel baseeruv täkkude paremusjärjestus ja kommentaarid hinnatud aretusväärtuste ja täkkude paremusjärjestuse sõltuvusest kasutatavast hindamismetoodikast.

Üks variant on presenteerida tulemused järgmise tabeli kujul:

Isa	Aretusväärtuste hinnangud ja täkkude paremusjärjestus					
	Selektsiooniindeks		Isa mudel		Looma mudel	
	Koht		Koht		Koht	
	$h^2 = 0,25$		$h^2 = 0,25$		$h^2 = 0,25$	$h^2 = 0,6$
Vigur 682 E						
Tukker 703 E						
Tiktora 697 E						
Rosett 600 E						
Rommik 710 E						

ⁱ Nii hobuste kui ka nende isade nimed tuleb asendada arvuliste koodidega alates ühest, seejuures peab numeratsioon olema genealoogiline, st isadele peavad vastama väiksemad numbrid.

NB! Et ülesande teises punktis on öeldud, et Tiktora 697 E on Tukker 703 E isa, siis oleks mõistlik juba isamudeli tarvis andmeid kodeerides omistada Tiktora 697 E-le väiksem arvuline kood.

ⁱⁱ Aretusväärtuste hindamine selektsiooniindeksite abil oli ülesandeks peatüki 5 lõpus. Kui seda tehtud pole, siis võite täkkude tõutüübi aretusväärtused nende järglaste kaudu hinnata ka nüüd. Seejuures tuleks rakendada järglaste arvu arvesse võtvat valemit

$$\hat{A}_i = \frac{\frac{1}{2}n_i h^2}{1 + \frac{1}{4}h^2(n_i - 1)} (\bar{P}_i - \bar{P}), \quad \text{kus } A_i \text{ on } i\text{-nda isa aretusväärtus, } n_i \text{ on } i\text{-nda isa järglaste arv, } \bar{P}_i \text{ on } i\text{-nda isa järglaste keskmine tõutüüp ja } \bar{P} \text{ on kõigi andmestikku kuuluvate hobuste keskmine tõutüüp.}$$