

ÜLEVAADE GENEETILISTE PARAMEETRITE HINDAMISEL KASUTATAVATEST MUDELITEST

Tanel Kaart

Sissejuhatus

Lihtsaim aretuslike probleemide lahendamisel kasutatav geneetiline mudel esitab indiviidi fenotüübiväärtuse P populatsiooni keskmise fenotüübiväärtuse \bar{P} , isendi genotüübiväärtuse G ja juhuslike keskkonnamõjude E summana:

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

Tänapäeva loomakasvatusteadus ei rakenda tavaliselt eeltoodud mudelit otse, vaid vaatlleb selle modifikatsioone, kus nii isendi genotüübiväärtus kui ka juhuslik keskkonnaefekt on omakorda esitatud erinevate mõjude summana. Viimaste hindamine võimaldab täpsustada looma aretusväärtuse hinnangut ning saada enam teavet uuritavat tunnust mõjutavate tegurite kohta. Mõningate selliste mudelite tutvustamine ongi üheks käesoleva kirjatüki ülesandeks.

Teiseks ülesandeks oleks esitada kasutatavad mudelid paralleelselt nii geneetika kui ka matemaatilise statistika terminoloogiat ja kirjapilti kasutades. Ilma geneetilist mudelit matemaatilise statistika keelde ümber panemata pole võimalik selle parameetreid statistika protseduuride abil hinnata. Mõistmata aga geneetika seaduspärasid on suur oht rakendada arvuti poolt väljastatud tulemusi vääralt.

Mudeleid võib grupeerida mitmel moel. Kõige üldisem oleks nende eristamine geneetilise varieeruvuse allika alusel. Esimese grupi moodustavad siin mudelid, kus hinnatakse vaid vanematelt sugurakkudega järglastele edasikanduvat geneetilist informatsiooni, s.t et hinnatakse vanemate aretusväärtusi järglaste baasil (pool- ja täisõvede analüüs). Teise grupi moodustavad nn looma mudelid, kus, nagu nimigi ütleb, hinnatakse mõõdetud loomade endi aretusväärtusi. Analüüsitav loom ise on ka geneetilise varieeruvuse allikaks. Lisaks tuleks eraldi vaadata emapoolset ja pesakonna (ühise keskkonna) mõju arvesse võtvaid mudeleid, samuti korduvmõõtmisi ning QTL-efektiga mudeleid. Peale selle on kõik mudelid esitatavad nii ühe- kui ka mitmemõõtmelisel kujul.

Poolõvede analüüs

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). Tavaliselt on selleks analüüsitavaks vanemaks isa (siit ka enamtuntud nimetus – isa mudel).

Et indiviidi genotüüpi kuuluvatest alleelidest on pooled pärit isalt ja pooled emalt, võib looma genotüübiväärtuse G esialgses mudelis asendada summaga $\frac{1}{2}A_S + \frac{1}{2}A_D + MS$ (tabel 1), kus A_S ja A_D on vastavalt isa ja ema aretusväärtused (alleeliväärtuste summad) ning MS väljendab Mendeli valiku mõju (Mendeli valikuks nimetatakse meioosi protsessis tehtavat lookusesisest alleelide vahelist valikut – valitakse, milline alleelidest järglasele pärandub). Alleelide dominantsusest tingitud efekt ja geenide mitteaditiivsest toimest tingitud epistaasi efekt loetakse kuuluvaks juhuslike keskkonnamõjude hulka, sest teatavasti päranduvad vanemaitl järglastele vaid üksikalleelid, mitte aga nendevahelised interaktsioonid, ning seega pole viimaste mõju järglaste järgi võimalik hinnata (Teinberg, 1978).

Matemaatilises (statistilises) mudelis esitatakse uuritava tunnuse väärtus y_{ij} (isendi j fenotüübiväärtus) vabaliikme μ (populatsiooni keskmine), isa i mõju s_i ja juhusliku vea e_{ij} summana (tabel 1). Seega on statistiliselt hinnatavad populatsiooni keskmine fenotüübiväärtus, isa mõju ning juhuslik viga, viimane sisaldab endas ka ema ja Mendeli valiku mõju.

Tabel 1. Poolõvede analüüsi geneetiline ja statistiline mudel ning vastavad dispersioonid

Table 1. Genetic and statistical model for halfsib analyses and associated covariance structure

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_{ij} = \bar{P} + \frac{1}{2}A_{S_i} + \frac{1}{2}A_{D_{ij}} + MS_{ij} + E_{ij},$ $\sigma_P^2 = \frac{1}{4}\sigma_A^2 + \left[\frac{1}{4}\sigma_A^2 + \frac{1}{2}\sigma_A^2 + \sigma_E^2 \right]$	$y_{ij} = \mu + s_i + e_{ij},$ $D \begin{pmatrix} \mathbf{s} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_a \sigma_s^2 & 0 \\ 0 & \mathbf{I}_n \sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_s^2 + \sigma_e^2$

($i = 1, \dots, a$, a - isade arv; $j = 1, \dots, n$, n - loomade arv andmestikus)

($i = 1, \dots, a$, a – number of sires; $j = 1, \dots, n$, n – number of animals)

Lugedes isa mõju juhuslikuks, avaldub uuritava tunnuse y dispersioon isast tingitud dispersiooni σ_s^2 ja juhuslikule veale vastava dispersiooni σ_e^2 summana. Seejuures eeldatakse siin keskkonna ja genotüübi sõltumatust, mis dispersioonimaatriksis on kirjas kujul $\text{Cov}(\mathbf{s}, \mathbf{e}') = 0$. Samuti eeldatakse vaatluste sõltumatust ja võrdset keskkonnatingimustest tingitud varieeruvust, s.t $D(\mathbf{e}) = \mathbf{I}_n \sigma_e^2$. Et $D(\mathbf{s}) = \mathbf{I}_a \sigma_s^2$, siis on kõigi geneetiliste efektide dispersioon σ_s^2 ja isade vaheline korrelatsioon on 0 (isad pole omavahel sugulased).

Geneetilises mudelis on fenotüübidispersioon lahutatav kaheks komponendiks – aditiivdispersiooniks σ_A^2 ja keskkonnadispersiooniks σ_E^2 . Lähtudes fenotüübiväärtuse esitusest võib fenotüübidispersiooni kirjutada välja ka pikemalt (vt tabel 1), millest selgub, et kogu geneetilisest varieeruvusest σ_A^2 on tervelt pool tingitud Mendeli valikust. Statistiliselt on poolõvede analüüsiga võimalik hinnata vaid neljandikku aditiivdispersioonist, s.o isa mõjust tingitud varieeruvust. Nii ema kui ka Mendeli valiku mõjust tingitud varieeruvus loetakse kuuluvaks keskkonnadispersiooni hulka.

Poolõvede analüüsist leitavaks geneetiliseks parameetrik on päritavuskoeffitsient (päritava genotüübilise muutlikkuse osakaal populatsiooni fenotüübilises muutlikkuses), mis arvutatakse suhtest

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2}.$$

Täisõvede analüüs

Juhul, kui indiviidi aretusväärtuse hinnang baseerub täisõvede analüüsil (analüüsitakse mõlema vanema aditiivset geneetilist mõju järglastele), on kasutatav mudel eelnevast pisut erinev. Peame 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 geneetilises mudelis kokku ühte, nn ühist keskkonna mõju iseloomustavasse liidetavasse E_C (tabel 2). Märgime veel, et Mendeli valiku mõju täisõvedele on erinev (kui see nii ei ole, on ilmselt tegu ühemunakaksikutega).

Statistilise mudeli kirjapanekul arvestame, et tavaliselt on üks isa paaritatud mitme emase loomaga ning tihti on andmete struktuur hierarhiline, s.t iga ema on paaritatud vaid ühe kindla isaga. Sellisel juhul on ema j mõju $d_{i(j)}$ allutatud isa i mõjule s_i .

Uuritava tunnuse dispersioon σ_y^2 avaldub isast tingitud dispersiooni σ_s^2 , emast tingitud dispersiooni σ_d^2 ja juhuslikule veale vastava dispersiooni σ_e^2 summana. Jällegi eeldatakse keskkonna ja genotüübi sõltumatust, vaatluste sõltumatust ning võrdset keskkonnast tingitud varieeruvust. Samuti on eelduseks emalt ja isalt päranduva geneetilise informatsiooni sõltumatus (võimalikku sugulust arvesse ei võeta).

Geneetilises mudelis on fenotüübidispersioon lahutatav kolmeks liidetavaks – aditiivdispersiooniks σ_A^2 , ühisest keskkonnamõjust tingitud varieeruvust kirjeldavaks komponendiks $\sigma_{E_C}^2$ ja keskkonnadispersiooniks σ_E^2 .

Tabel 2. Täisõvede analüüsi geneetiline ja statistiline mudel ning vastavad dispersioonid

Table 2. Genetic and statistical model for fullsib analyses and associated covariance structure

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_{ijk} = \bar{P} + \frac{1}{2}A_{S_i} + \frac{1}{2}A_{D_j} + MS_{ijk} + E_{C_{ij}} + E_{ijk},$ $\sigma_P^2 = \frac{1}{4}\sigma_A^2 + \left[\frac{1}{4}\sigma_A^2 + \sigma_{E_c}^2\right] + \left[\frac{1}{2}\sigma_A^2 + \sigma_E^2\right]$	$y_{ijk} = \mu + s_i + d_{j(i)} + e_{ijk},$ $D \begin{pmatrix} \mathbf{s} \\ \mathbf{d} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_a \sigma_s^2 & 0 & 0 \\ 0 & \mathbf{I}_b \sigma_d^2 & 0 \\ 0 & 0 & \mathbf{I}_n \sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_s^2 + \sigma_d^2 + \sigma_e^2$

($i = 1, \dots, a$, a - isade arv; $j = 1, \dots, b_i$, $b_i - i$. isaga paaritatud emaste loomade arv; $k = 1, \dots, n$, n - loomade arv andmestikus)

($i = 1, \dots, a$, a - number of sires; $j = 1, \dots, b_i$, $b_i - i$ - number of dams mated with i -th sire; $k = 1, \dots, n$, n - number of animals)

Sarnaselt isa mudelile on liidetavad grupeeritavad (vt tabel 2). Kogu geneetilisest varieeruvusest σ_A^2 on jällegi tervelt pool tingitud Mendeli valikust ning statistilisel analüüsil loetakse see kuuluvaks keskkonnadispersiooni hulka. Ülejäänud poolest on vaid isalt pärandunud osa hinnatav nihketa, ema mõju on seotud ühise keskkonnamõjuga ja seetõttu pole ka statistilises mudelis võimalik eraldada emast tingitud geneetilist varieeruvust ühisest keskkonnamõjust tingitud varieeruvusest.

Täisõvede analüüsil on leitavaid geneetilisi parameetreid kolm. Kõik nad arvutatakse geneetilisest mudelist leiduvate dispersioonikomponentide suhetena kogu varieeruvusse:

päritavus

$$h^2 := \frac{\sigma_A^2}{\sigma_P^2} = \frac{4\sigma_s^2}{\sigma_y^2}, \text{ kus}$$

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 ($\sigma_d^2 = \frac{1}{4}\sigma_A^2 + \sigma_{E_c}^2$, vt tabel 2);

ühisest keskkonna mõjust tingitud varieeruvuse osakaal koguarveeruvuses

$$c^2 := \frac{\sigma_{E_c}^2}{\sigma_p^2} = \frac{\sigma_d^2 - \sigma_s^2}{\sigma_y^2};$$

juhuslikust keskkonna mõjust tingitud varieeruvuse osakaal koguarveeruvuses

$$e^2 := \frac{\sigma_E^2}{\sigma_p^2} = \frac{\sigma_e^2 - 2\sigma_s^2}{\sigma_y^2}.$$

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

Looma mudel

Looma mudeli abil hinnatakse vaadeldud loomade endi aretusväärsusi. Loomadevahelist sugulust arvesse võttes saame leida aretusväärtused ka ilma vaatlustulemusteta isendeile (näiteks pullidele piimatoodangu või jääradele poegimisraskuse osas).

Geneetilisest mudelist avaldub uuritava tunnuse fenotüübiväärtus populatsiooni keskmise, looma aretusväärtuse ja juhusliku vea summana. Samadest liidetavatest koosneb ka statistiline mudel.

Uuritava tunnuse dispersioon σ_y^2 avaldub aditiivdispersiooni σ_a^2 ja juhuslikule veale vastava dispersiooni σ_e^2 summana. Sarnaselt isa mudeliga eeldatakse keskkonna ja genotüübi sõltumatust, vaatluste sõltumatust ning võrdset keskkonnast tingitud varieeruvust. Küll aga võetakse arvesse võimalik loomadevaheline sugulus.

Fakt, et kaks looma on sugulased, ilmneb selles, et nendevaheline geneetilisest faktoritest tingitud kovariatsioon on nullist erinev. Olulisim 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. Aditiivne geneetiline sugulus loomade vahel on väljendatav aditiivse geneetilise suguluse maatriksi \mathbf{A} abil (Henderson, 1984). Maatriksi \mathbf{A} on $n \times n$ maatriks, kus n on uuritavate loomade arv. Maatriksi \mathbf{A} elemendid a_{ij} on võrdsed kahekordse tõenäosusega, et juhuslik alleel indiviidi i genotüübist on identne päritolult vastava alleeliga indiviidi j genotüübist, st. et peadiagonaalil paiknevate elementide (loomade sugulus iseendaga) väärtuseks on 1 pluss looma inbriidingukoefitsient, väljaspool peadiagonaali paiknevad aga loomadevahelised suguluskoefitsiendid (mis on $1/2$ täisõvede või vanema-järglase vahel, $1/4$ poolõvede või looma ja tema vanavanemate vahel jne.).

Esitades loomadevahelist aditiivgeneetilist varieeruvust kirjeldava maatriksi kujul $\mathbf{D}(\mathbf{a}) = \mathbf{A}\sigma_a^2$, saame statistilisel analüüsil arvesse võtta ka

inbriidingut ja selektsiooni (Henderson, 1984; van Arendonk, Van der Werf, 1997).

Tabel 3. Looma mudel

Table 3. Animal model

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_i = \bar{P} + A_i + E_i,$ $\sigma_P^2 = \sigma_A^2 + \sigma_E^2$	$y_i = \mu + a_i + e_i,$ $D \begin{pmatrix} \mathbf{a} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & 0 \\ 0 & \mathbf{I}_n\sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_a^2 + \sigma_e^2$

($i = 1, \dots, n$, n - loomade arv andmestikus)

($i = 1, \dots, n$, n - number of animals)

Hinnatavaid geneetilisi parameetreid on vaid üks, s.o päritavuskoeffitsient

$$h^2 := \frac{\sigma_A^2}{\sigma_P^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.$$

Korduvmõõtmised

Korduvate mõõtmiste (ühte sama tunnust on ühel ja samal isendil mõõdetud mitu korda) korral avaldub looma fenotüübiväärtus P tema genotüübiväärtuse G , keskkonna püsiva (permanentse) mõju E_p ja juhusliku (temporaalse) keskkonna mõju E_t summana.

Statistilises mudelis on looma i mõju kajastavas liidetavas g_i summeeritud nii looma genotüübiväärtus kui ka keskkonna permanentne mõju (tabel 4).

Et korduvmõõtmiste korral pole võimalik teineteisest eraldada geneetilist ja permanentset keskkonna mõju ega ka neist teguritest tingitud varieeruvust, on hinnatavaid geneetilisi parameetrid vaid üks, s.o korduvus,

$$r := \frac{\sigma_G^2 + \sigma_{E_p}^2}{\sigma_P^2} = \frac{\sigma_g^2}{\sigma_y^2}.$$

Tabel 4. Korduvmõõtmistega analüüsi geneetiline ja statistiline mudel ning vastavad dispersioonid

Table 4. Model with repeated measurements and associated covariance structure

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_{ij} = \bar{P} + G_i + E_{p_i} + E_{t_{ij}},$ $\sigma_P^2 = [\sigma_G^2 + \sigma_{E_p}^2] + \sigma_{E_t}^2$	$y_{ij} = \mu + g_i + e_{ij},$ $D \begin{pmatrix} \mathbf{g} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n \sigma_g^2 & 0 \\ 0 & \mathbf{I}_N \sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_g^2 + \sigma_e^2$

($i = 1, \dots, n$, n - erinevate loomade arv;

$j = 1, \dots, N$, N - andmestiku suurus)

($i = 1, \dots, n$, n – number of animals;

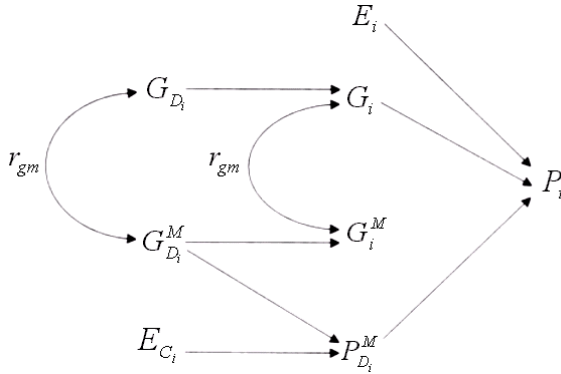
$j = 1, \dots, N$, N – number of observations in dataset)

Korduvust võib määratleda kui fenotüübilise dispersiooni seda osa, mis on ühine antud tunnuse korduval mõõtmisel, samuti annab korduvus vaadeldava tunnuse päritavuse ülemise piiri (kui $\sigma_{E_p}^2 \rightarrow 0$ ja $\sigma_G^2 \rightarrow \sigma_A^2$, siis $r \rightarrow h^2$).

Emapoolne looma mudel

Paljudel juhtudel ei rahulda tänapäeva aretajaid ja uurijaid enam eelnevad küllalt lihtsad mudelid. Aretusprotsessi kiirendamise, generatsiooniintervalli vähendamise ja raha kokkuhoiu huvides on vajalik teostada selektsioon loomade hulgas nii varakult kui võimalik. Noorloomadel mõõdetavad tunnused sõltuvad aga üsna palju nende emast ja/või neid ümbritsevast keskkonnast (pesakonna mõju). Emapoolselt mõjutatud tunnuste hulka kuuluvad näiteks lihaveiste võõrutusmass, põrsaste pesakonnamass jne.

Ema mõju korral on indiviidi i fenotüübiväärtus P_i esitatav kahe komponendi summana. Neist esimene, nn otsene fenotüübiväärtus P_i^0 , võrdub populatsiooni keskmise fenotüübiväärtuse \bar{P} , isendi genotüübiväärtuse G_i ja talle mõjuvate juhuslike keskkonnamõjude E_i summana: $P_i^0 = \bar{P} + G_i + E_i$ (joonis 1). Teine komponent väljendab mitteotsest ema mõju ja jaguneb samuti geneetiliseks ja mittegeneetiliseks osaks: $P_{D_i}^M = G_{D_i}^M + E_{C_i}$, kus D_i märgib looma i ema, $G_{D_i}^M$ tähistab ema geneetilist võimet hoolitseda oma järglas(t)e eest ja E_{C_i} märgib loomale i mõjuvaid oma ema teiste järglastega või pesakonnakaaslastega ühiseid keskkonna mõjusid (Willham, 1972; Lynch, Walsh, 1998).



Joonis 1. Rajaskeem, mis illustreerib indiviidi i fenotüübiväärtuse P_i sõltuvust otsesest geneetilisest efektist G_i , juhuslikest keskkonnamõjudest E_i , emapoolsest geneetilisest efektist $G_{D_i}^M$ ja ühistest keskkonnamõjudest E_{C_i} ; r_{gm} on otseste ja emapoolsete efektide vaheline geneetiline korrelatsioon

Figure 1. Path diagram representing the determination of the phenotype P_i of an individual i by direct genetic effects G_i , direct environmental effects E_i , maternal genetic effects $G_{D_i}^M$ and common environmental effects E_{C_i} ; r_{gm} is the genetic correlation between direct and maternal effects.

Et prognoosiv on jällegi vaid geenide aditiivne efekt, asendatakse mudelis looma genotüübiväärtus G_i tema aretusväärtusega A_i ja ema mitteotsene geneetiline mõju $G_{D_i}^M$ tema aditiivgeneetilise mõjuga M_{D_i} (tabel 5).

Samad liidetavad, μ – populatsiooni keskmine, a_i – looma i aretusväärtus, m_j – ema j mõju, c_k – pesakonna k mõju ja e_{ijk} – juhuslik viga, on hinnatavad ka statistilisest mudelist.

On loomulik oletada, et otsesed ja emapoolsed geneetilised efektid ei ole sõltumatud – samad geenid, mis määravad ära emise kehamassi võivad mõjutada ka tema piimakust. Seetõttu eeldatakse, et leidub korrelatsioon r_{am} otseste ja emapoolsete aditiivgeneetiliste efektide vahel. Sarnaselt eelnevatele mudelitele eeldatakse, et keskkond ja genotüüp on sõltumatud. Küll aga võetakse arvesse loomadevaheline sugulus ja seda nii aretusväärtuste kui ka emapoolsete efektide prognoosimisel (sel teel saab näiteks leida, kui head emaomadused pärandab jäär või kult oma järglastele). Statistilises mudelis on loomadevaheline sugulus kirjas aditiivse geneetilise suguluse maatriksi **A** kujul.

Tabel 5. Emapoolne looma mudel**Table 5.** Maternal animal model

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_i = \bar{P} + A_i + M_{D_i} + E_{C_i} + E_i,$ $\sigma_P^2 = \sigma_A^2 + \sigma_M^2 + 2Cov(A, M) + \sigma_{E_c}^2 + \sigma_E^2$	$y_{ijk} = \mu + a_i + m_j + c_k + e_{ijk},$ $D \begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{c} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & 0 & 0 \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & 0 & 0 \\ 0 & 0 & \mathbf{I}_b\sigma_c^2 & 0 \\ 0 & 0 & 0 & \mathbf{I}_n\sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_a^2 + \sigma_m^2 + 2\sigma_{am} + \sigma_c^2 + \sigma_e^2$

($i, j = 1, \dots, n, n$ - loomade arv andmestikus;

$k = 1, \dots, b, b$ - erinevate pesakondade/emade arv)

($i, j = 1, \dots, n, n$ - number of animals;

$k = 1, \dots, b, b$ - number of litters/mothers)

Emapoolsest loomamudelist on leitav terve hulk geneetilisi parameetreid, esmalt muidugi päritavus

$$h^2 := \frac{\sigma_A^2}{\sigma_P^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},$$

lisaks ka emapoolsest mitteotsesest mõjust tingitud varieeruvuse osakaal koguvarieeruvuses

$$m^2 := \frac{\sigma_M^2}{\sigma_P^2} = \frac{\sigma_m^2}{\sigma_y^2},$$

ühisest keskkonna mõjust tingitud varieeruvuse osakaal koguvarieeruvuses

$$c^2 := \frac{\sigma_{E_c}^2}{\sigma_P^2} = \frac{\sigma_c^2}{\sigma_y^2},$$

otseste ja emapoolsete aditiivgeneetiliste efektide vaheline korrelatsioon

$$r_{am} := \frac{Cov(A, M)}{\sqrt{\sigma_A^2 \cdot \sigma_M^2}} = \frac{\sigma_{am}}{\sqrt{\sigma_a^2 \cdot \sigma_m^2}}.$$

QTL-efektiga mudelid

Kõigi eelnevate mudelite korral eeldati, et uuritavate tunnuste väärtused on määratud lõpmatu paljude lõpmatu väikese mõjuga geenide poolt. Ometi on viimastel aastatel tuvastatud rida genee, mis määravad üksi ära suure osa kvantitatiivgeneetiliselt päranduva tunnuse potentsiaalsest väärtusest. QTL-i (ingl. quantitative trait loci) all mõistetaksegi tänapäeval lookust, milles paiknev geen omab uuritavale fenotübile märkimisväärset mõju (QTL-efekt on

siis QTL-is paikneva geeni mõju). Sellisel juhul tuleb loomade aretusväärtuste hindamisel lisaks tavapärasele polügeensele efektile arvesse võtta ka QTL-i(de) mõju(d) (Kinghorn jt., 2000).

Vastava mudeli lihtsaim variant on toodud tabelis 6. Looma aretusväärtus on esitatav kahe komponendi, aditiivgeneetilise mõju ja QTL-i mõju, summana (ühe QTL-i asemel võib märkimisväärse mõjuga geene olla ka rohkem – mudelisse tuleb lihtsalt rohkem liikmeid). QTL-efekt ise on ka esitatav kahe komponendi summana, need on vastavalt emalt ja isalt päritud QTL-s paiknevate alleelide efektid (q_i^f ja q_i^m). Loomade valikut sellise mudeli alusel nimetatakse markerseleksiooniks (ingl. marker assisted selection – MAS).

QTL-efekti reaalseks hindamiseks on tänapäeval vajalik QTL-ga aheldunud markerite olemasolu. Just geneetiliste markerite kohta teadaoleva info ja eelnevalt fikseeritud markerite ja QTL-i vaheliste rekombinatsioonide esinemise tõenäosuse alusel arvutatakse gameetilise suguluse maatriksi \mathbf{Q} elemendid, milledeks on tõenäosused, et QTL-s paiknevad alleelid on identsed päritolult (Fernando, Grossman, 1989; van Arendonk jt., 1994; Wang jt., 1995). Gameetilise suguluse maatriksi dimensioon on $2n \times 2n$, kus n on uuritavate loomade arv ja $2n$ vaatluse all olevate alleelide arv (igal loomal on ühest geenist kaks alleeli). Loomadevaheline aditiivgeneetiline sugulus võetakse arvesse sarnaselt looma mudelile aditiivse geneetilise suguluse maatriksi \mathbf{A} läbi. Eeldatakse, et puudub korrelatsioon keskkonna ja geenide vahel, samuti paljude väikese mõjuga geenide summaarse mõju ja QTL-i vahel. Uuritava tunnuse dispersioon on siis esitatav kolme komponendi summana, kus esimesed kaks, σ_a^2 ja σ_q^2 (vastavalt aditiivdispersioon ja QTL-efektist põhjustatud varieeruvus),

Tabel 6. QTL-efektiga looma mudel
Table 6. Animal model with QTL-effect

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_i = \bar{P} + A_i + Q_i + E_i,$ $\sigma_p^2 = \sigma_A^2 + \sigma_Q^2 + \sigma_E^2$	$y_i = \mu + a_i + q_i^m + q_i^f + e_i,$ $D \begin{pmatrix} \mathbf{a} \\ \mathbf{q} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & 0 & 0 \\ 0 & \mathbf{Q}\sigma_q^2 & 0 \\ 0 & 0 & \mathbf{I}_n\sigma_e^2 \end{pmatrix},$ $\sigma_y^2 = \sigma_a^2 + \sigma_q^2 + \sigma_e^2$

($i = 1, \dots, n$, n - loomade arv andmestikus)

($i = 1, \dots, n$, n – number of animals)

väljendavad tegelikult mõlemad geenidest tingitud varieeruvust loomade vahel ja vaid kolmas liidetav, σ_e^2 , mõõdab keskkonnatingimustest põhjustatud erinevuse suurust.

Olulisim hinnatav geneetiline parameeter on jällegi päritavuskoeffitsient

$$h^2 := \frac{\sigma_A^2 + \sigma_Q^2}{\sigma_P^2} = \frac{\sigma_a^2 + \sigma_q^2}{\sigma_a^2 + \sigma_q^2 + \sigma_e^2}.$$

Vastavate dispersioonikomponentide suhetena koguvarieeruvusse on leitavad ka aditiivgeneetilise varieeruvuse ja QTL-efektist põhjustatud varieeruvuse osakaalud.

Mitmemõõtmelised mudelid

Et põllumajandusloomade seleksioon toimub tavaliselt mitme tunnuse alusel, on korrektse tulemuse tarvis teada ka nende tunnuste vahelisi seoseid (nii fenotüübilisi kui ka geneetilisi). Tunnustevahelisi geneetilisi seoseid, mida tavaliselt põhjendatakse pleiotroopsuse efektiga (üks geen mõjutab korraga mitut tunnust), mõõdetakse geneetilise korrelatsiooni abil. Tunnuste vaheliste korrelatsioonide arvestamine suurendab hinnatavate parameetrite (aretusväärtused, emapoolsed ja pesakonna efektid) täpsust ja võimaldab loomi õigustatumalt selekteerida (Schaeffer, 1984; Thompson, Meyer, 1986). Mitmemõõtmeline analüüs on ka ainus võimalus saada nihketa hinnangud tunnusele, mida mõõdetakse eelnevalt juba teise, mõõdetava tunnusega korreleeruva tunnuse alusel selekteeritud loomadel. Näide selle kohta on veiste teise laktatsiooni jõudluse hindamine, kus eelnev seleksioon loomade hulgas on tehtud esimese laktatsiooni toodangute alusel, kirjed teise laktatsiooni toodangunäitajatega on vaid esimese laktatsiooni järel karja alles jäetud lehmadel. Esimese laktatsiooni näitajaid arvestamata saame teise laktatsiooni toodangute alusel leitud populatsiooni geneetilistele parameetritele nihkega hinnangud (Mrode, 1996).

Lihtsaima näitena võiks vaadelda kahemõõtmelist looma mudelit (tabel 7). Mõlema uuritava tunnuse fenotüübiväärtus avaldub populatsiooni keskmise, looma aretusväärtuse ja juhusliku vea summana (kumma tunnusega on tegu, näitab indeks j).

Tunnuste varieeruvust kirjeldava dispersioonimaatriksi dimensioon on 2 korda kõrgem, kui ühemõõtmelise mudeli korral – lisandunud on aretusväärtuste vaheline kovariatsioon $\sigma_{a_1 a_2}$ ja keskkonnamõjude vahelist seost iseloomustav liige $\sigma_{e_1 e_2}$.

Nagu ühemõõtmelisest looma mudelist saab siitki hinnata uuritavate tunnuste päritavust (indeks näitab jällegi, kumma tunnusega on tegu):

$$h_1^2 := \frac{\sigma_{A_1}^2}{\sigma_{P_1}^2} = \frac{\sigma_{a_1}^2}{\sigma_{a_1}^2 + \sigma_{e_1}^2}, \quad h_2^2 := \frac{\sigma_{A_2}^2}{\sigma_{P_2}^2} = \frac{\sigma_{a_2}^2}{\sigma_{a_2}^2 + \sigma_{e_2}^2};$$

Tabel 7. Kahemõõtmeline looma mudel

Table 7. Two-dimensional animal model

Geneetiline mudel Genetic model	Statistiline mudel Statistical model
$P_{ij} = \bar{P}_j + A_{ij} + E_{ij},$ $\sigma_{P_j}^2 = \sigma_{A_j}^2 + \sigma_{E_j}^2$	$y_{ij} = \mu_j + a_{ij} + e_{ij},$ $D \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_{a_1}^2 & \mathbf{A}\sigma_{a_1a_2} & 0 & 0 \\ \mathbf{A}\sigma_{a_1a_2} & \mathbf{A}\sigma_{a_2}^2 & 0 & 0 \\ 0 & 0 & \mathbf{I}_n\sigma_{e_1}^2 & \mathbf{I}_n\sigma_{e_1e_2} \\ 0 & 0 & \mathbf{I}_n\sigma_{e_1e_2} & \mathbf{I}_n\sigma_{e_2}^2 \end{pmatrix},$ $\sigma_{y_j}^2 = \sigma_{a_j}^2 + \sigma_{e_j}^2$

($i = 1, \dots, n$, n - loomade arv andmestikus; $j = 1, 2$ (2 tunnust))

($i = 1, \dots, n$, n - number of animals; $j = 1, 2$ (2 traits))

lisaks leitakse ka kõikvõimalikud tunnuste vahelised korrelatsioonid:
geneetiline korrelatsioon

$$r_g := \frac{Cov(A_1, A_2)}{\sqrt{\sigma_{A_1}^2 \cdot \sigma_{A_2}^2}} = \frac{\sigma_{a_1a_2}}{\sqrt{\sigma_{a_1}^2 \cdot \sigma_{a_2}^2}};$$

keskkonnast tingitud korrelatsioon

$$r_e := \frac{Cov(E_1, E_2)}{\sqrt{\sigma_{E_1}^2 \cdot \sigma_{E_2}^2}} = \frac{\sigma_{e_1e_2}}{\sqrt{\sigma_{e_1}^2 \cdot \sigma_{e_2}^2}};$$

fenotüübiline korrelatsioon

$$r_p := \frac{Cov(P_1, P_2)}{\sqrt{\sigma_{P_1}^2 \cdot \sigma_{P_2}^2}} = \frac{\sigma_{y_1y_2}}{\sqrt{\sigma_{y_1}^2 \cdot \sigma_{y_2}^2}}.$$

Analoogselt on konstrueeritavad ka keerulisemad ja enam kui kahemõõtmelised mudelid. Eestis rakendati mitmemõõtmelist looma mudelit esmakordselt 1996. aastal veiste piimajõudluse aretusväärtuste hindamisel (Reents jt., 1996). Mitmemõõtmeline emapoolse efektiga loomamudel sai praktilise väljundi alles 2000. aastal, mil seda kasutati lammaste jõudlusandmete analüüsil (Kaart ja Piirsalu, 2000).

Diskussioon

Siin toodud mudelid on vaid kõige olulisemad ja üldisemad. Kõigist neist leidub rida modifikatsioone. Näiteks leiab sageli kasutust sugulusega isa mudel. Selles võetakse arvesse võimalik sugulus looma isa ja ema isa vahel (võimalik ema efekt loetakse endiselt mitteoluliseks, sest tavaliselt on see segatud

keskkonnamõjudega, samuti on üldjuhul järglasi ühe ema kohta märksa vähem kui ühe isa kohta, mis muudab ema mõju hinnangu võrreldes isa mõjuga ebatäpsemaks). Sel teel saab leida aretusväärtused ka vaatlustulemusega järglast mitteomavatele isadele.

Täiesti jäi vaatluse alt välja nn kitsendatud looma mudel (Quaas, Pollak, 1980), mis kombineerib omavahel isa ja looma mudeli, käsitledes vanemaid looma mudeli ja järglasi isa mudeli alusel. Mudelit rakendatakse väga suurte andmemassiivide, mille korral tavalise loomamudeli kasutamiseks napib ressursse, analüüsimiseks. Hinnata võimaldab see mudel vaid vanemate aretusväärtusi.

Et päritavuskoefitsient on hinnatav kõigist mudelitest (va kordumõõtmiste korral), tekib koheselt ka küsimus, millise mudeli korral on hinnang õigeim. Ega siin ühest vastust pole. Arutleda võib vaid selle üle, kui suur osa tegelikust päritavast genotüübilisest muutlikkusest ühe või teise mudeliga „kinni püütakse”. Isa ja looma mudelist hinnatud päritavuskoefitsientide sarnasus või erinevus sõltub eelkõige andmete struktuurist, ilmne on vaid, et looma mudelist saadud päritavuskoefitsiendi hinnang on täpsem (ja enamikul juhtudest ka pisut suurem). Võrreldes looma ja emapoolse efektiga looma mudelit on selge, et emapoolse mõjuga looma mudeli korral on aditiivgeneetiline varieeruvus ja seega siis ka päritavuskoefitsiendi väärtus väiksemad (osa looma mudelis aditiivgeneetilise dispersiooni hulka arvatud varieeruvusest kirjeldab tegelikult emapoolsest mitteotsesest geneetilisest mõjust tingitud muutlikkust). Päritava genotüübilise muutlikkuse osakaal (päritavus) on suurim QTL-efektiga mudelis, sest selles lisandub tavapärasele geenide aditiivsele mõjule ka QTL-s paikneva märkimisväärse mõjuga geeni efekt.

Kirjandus

1. Fernando, R.L., Grossman, M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics, Selection, Evolution*, **21**, 467-477.
2. Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph, Canada.
3. Kaart, T., Piirsalu, P. (2000) The complex analysis of genetic parameters in Estonian sheep breeds. *Proceedings of the 6th Baltic Animal Breeding Conference, 27-28 April, 2000, Jelgava*, 135-140.
4. Kinghorn, B.P., Van der Werf, J., Ryan, M. (2000) *Animal Breeding: Use of New Technologies*. Post Graduate Foundation in Veterinary Science, University of Sydney.
5. Lynch, M., Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer Associates, Inc.

6. Mrode, R.A. (1987) *Linear Models for the Prediction of Animal Breeding Values*. CAB International, UK.
7. Quaas, R.L., Pollak, E.J. (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *Journal of Animal Science*, **51**, 1277-1287.
8. Reents, R., Uba, M., Pedastsaar, K., Vares, T. (1996) Implementation of animal model for production traits of dairy cattle in Estonia. *Proceedings of the Open Session of the Interbull Annual Meeting, June 23-24, 1996, Veldhoven, The Netherlands*, Bulletin no. **14**, 135-139.
9. Schaeffer, L.R. (1984) Sire and cow evaluation under multiple trait models. *Journal of Dairy Science*, **67**, 1567-1580.
10. Teinberg, R. (1978) *Põllumajandusloomade geneetika*. "Valgus", Tallinn.
11. Thompson, R., Meyer, K. (1986) A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science*, **15**, 299-313.
12. van Arendonk, J.A.M., Van der Werf, J. (1997) *Breeding value estimation – lecture notes for E250-209*. Wageningen Agricultural University, Department of Animal Breeding.
13. van Arendonk, J.A.M., Tier, B., Kinghorn, B.P. (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics*, **137(1)**, 319-329.
14. Wang, T., Fernando, R.L., Van der Beek, S., Grossman, M., van Arendonk, J.A.M. (1995) Covariance between relatives for a marked quantitative trait locus. *Genetics, Selection, Evolution*, **27**, 251-274.
15. Willham, R. L. (1972) The role of maternal effects in animal breeding. III. Biometrical aspects of maternal effects in animals. *Journal of Animal Science*, **35**, 1288-1293.

OVERVIEW ABOUT THE MODELS USED ON THE ESTIMATION OF GENETIC PARAMETERS

Tanel Kaart

Summary

The simplest genetic model used in animal breeding present the phenotypic value of animal P as sum of mean \bar{P} , genetic value G and environmental value E . Nowadays animal science is using many different modifications of this model. The themes like half and full sib models, animal models, models with repeated measurements, maternal models, models with QTL-effect and multivariate models are considered. Concurrently the genetic and statistical models are presented and their differences discussed for different analyses. The estimable effects and genetic parameters are presented.