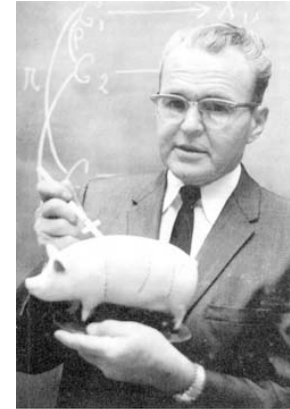


V

SELEKTSIOONIINDEKSID

Kuigi geneetiliste parameetrite (päritavuskoefitsiendid, geneetilised korrelatsioonikordajad, aretusväärtused) hindamiseks reaalses, suurtes ja väga erinevatel sugulusastmetel indiviididest koosnevas populatsioonis on nii tulemuste täpsuse kui ka arvutuste lihtsama teostatavuse huvides otstarbekam kasutada järgnevais punktides käsitletavaid üldistel lineaarsetel mudelitel baseeruvaid meetodeid, on loomade aretusstrateegiate välja töötamisel ja aretusprogrammide koostamisel aluseks ikkagi kindlatele põlvnemisskeemidele ja populatsioonigeneetika seaduspärasele tuginevad selektsiooniindeksid.

Esmalt (1936, Fairfield-Smith) taimekasvatuse tarvis välja töötatud selektsiooniindeksite teooriat arendas edasi ja kohandas loomade aretusele USA loomakasvatusteadlane **Lanoy Nelson Hazel** 1941. aastal oma doktoriväitkirjas. Laiemalt tuntuks sai see erinevate informatsiooniallikate ja andmete ühte mudelisse inkorporeerimist selgitav ning siimaani kogu kunstliku valiku aluseks olev teooria 1943. aastal peale publitseerimist ajakirjas „Genetics” (Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics, 28, 476-490). Hazel ise oli II Maailmasõja järgses USA-s (ja seeläbi ka maailmas) juhtivaks persooniks loomade aretusteooria väljatöötamisel. Muuhulgas kuulus tema poolt juhendatavate üliõpilaste hulka ka 1950.-ndatest alates loomade tõuaretuses uusi suundi rajanud ning vähimruutude meetodi ja selektsiooniindeksite teooria ühendamise läbi dispersioonanalüüsi segamudeliteni (*BLUP*, looma mudel jmt) jõudnud Charles Roy Henderson.



The Texas Youth

5.1 DEFINITSIOON

5.1.1 Seleksiooniindeks ühele tunnusele

Juhul kui kogu selektsioonialune populatsioon paikneb sarnastes keskkonnatingimustes, piisab geneetiliselt parimate indiviidide välja valimiseks iga indiviidi (ja/või tema sugulaste) fenotüübiväärtuste võrdlemisest populatsiooni keskmisega¹. Kõik need fenotüübil mõõdetud erinevused koondatakse sobivalt valitud kordajatega kaalutuna ühte võrrandisse. Sellist looma aretusväärtuse (või geneetilise väärtuse) määramiseks konstrueeritud võrrandit nimetatakse **selektsiooniindeksiks**.

Selektsiooniindeksi üldkuju on

$$I = b_1X_1 + b_2X_2 + \dots + b_mX_m, \quad (5.1)$$

kus X_i tähistab indiviidi enese või tema sugulase fenotüübiväärtuse (või fenotüübiväärtuste keskmise) erinevust populatsiooni keskmisest ja b_i on sobivalt valitud kaaluparameeter (mis vastavalt regressioonikordaja olemusele näitab muutust indeksi väärtuses fenotüübiväärtuse muutumisel ühe ühiku võrra). Parameetrid b_i püütakse valida nii, et indeks I kujutaks enesest parimat prognoosi aretusväärtusele A , $\hat{A} = I$. Parim tähendab siin, et

- korrelatsioon prognoositud aretusväärtuse (indeksi väärtuse) ja tegeliku aretusväärtuse vahel, $r_{\hat{A},A} = r_{I,A}$, on maksimaalne,
- prognooside ruutviga $(A - I)^2$ ja prognooside varieeruvus $\text{var}(I)$ on minimaalsed,
- indiviidide korrektne järjestamine nende geneetilise potentsiaali alusel toimub suurima tõenäosusega ning
- selektsiooni läbi saavutatav geneetiline edu on maksimaalne.

¹ Juhul, kui keskkonnatingimused on erinevad, aga nende mõju on täpselt teada, korrigeeritakse kõiki võrreldavaid suurusi keskkonnatingimuste suhtes (näiteks lahutatakse iga lehma piimatoodangust temale vastava farmi keskmise piimatoodangu erinevus populatsiooni keskmisest ja võrreldakse tulemust kõigi populatsiooni kuuluvate lehmade farmi mõju suhtes korrigeeritud piimatoodangute keskmisega).

Samuti teisendatakse kordajate b_i avaldisi nii, et need ei sõltuks uuritava tunnuse väärtustest konkreetse populatsioonis vaid avalduksid standardsete, eeldatavalt suhteliselt muutumatute ja normeeritud väärtustega populatsioonigeneetika parameetrite nagu päritavuskoefitsient, korduvus, geneetilised korrelatsioonikordajad jne funktsioonina.

Et selektsooniindeks on fenotüübil mõõdetud väärtuste suhtes lineaarne võrrand, nimetatakse teda ka **aretusväärtuse parimaks lineaarseks prognoosiks** (*best linear predictor, BLP*).

Ühe indiviidi ühele tunnusele konstrueeritud selektsooniindeks (5.1) on maatrikskujul järgmine:

$$I = \mathbf{b}^T \mathbf{X},$$

kus $\mathbf{b} = (b_1 \ b_2 \ \dots \ b_m)^T$ on selektsooniindeksi kaaluparameetrite vektor ja $\mathbf{X} = (X_1 \ X_2 \ \dots \ X_m)^T$ on fenotüübil mõõdetud erinevuste vektor.

Et selektsooniindeks on oma kujult mitmene regressioonivõrrand, on tema kordajad b_i avaldatavad seosest (vt pt 2)

$$b_i = \text{cov}(X_i, A) / \text{var}(X_i). \quad (5.2)$$

Sama valem maatrikskujul kirjutanduna:

$$\mathbf{b} = \text{var}(\mathbf{X})^{-1} \text{cov}(\mathbf{X}, A) = \mathbf{P}^{-1} \mathbf{G}, \quad (5.3)$$

kus $\text{var}(\mathbf{X}) = \mathbf{P}$ on fenotüübil mõõdetud erinevuste dispersioonimaatriks (dimensiooniga $m \times m$) ja $\text{cov}(\mathbf{X}, A) = \mathbf{G}$ on fenotüübi ja tegeliku genotüübi vaheliste kovariatsioonide $m \times 1$ -vektor²:

$$\mathbf{P} = \text{var}(\mathbf{X}) = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_m) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \dots & \text{cov}(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_m, X_1) & \text{cov}(X_m, X_2) & \dots & \text{var}(X_m) \end{pmatrix} \quad \text{ja} \quad \mathbf{G} = \text{cov}(\mathbf{X}, A) = \begin{pmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \\ \vdots \\ \text{cov}(X_m, A) \end{pmatrix}.$$

5.1.2 Selektsooniindeksi täpsus

Selektsooniindeksi kujul avalduva aretusväärtuse hinnangu usaldusväärsuse mõõtmisel lähtutakse hinnangu varieeruvusest ning hinnangu ja hinnatava parameetri tegeliku väärtuse sarnasusest.

Aretusväärtuse hinnangu varieeruvust mõõdavad **selektsooniindeksi dispersioon** $\text{var}(I) = \sigma_I^2$ ja standardhälve (viimane kujutab enesest hinnatava **aretusväärtuse standardviga**: $\text{se}(A) = \sigma_{\hat{A}} = \sigma_I$).

Lähtudes selektsooniindeksi (5.1) maatrikskujust $I = \mathbf{b}^T \mathbf{X}$ ja indeksi kordajate avaldisest (5.3), on indeksi dispersioon avaldatav maatriksvõrdusena

$$\sigma_I^2 = \text{var}(\mathbf{b}^T \mathbf{X}) = \mathbf{b}^T \text{var}(\mathbf{X}) \mathbf{b} = \mathbf{b}^T \mathbf{P} \mathbf{b} \stackrel{(5.3) \Rightarrow \mathbf{P} \mathbf{b} = \mathbf{G}}{=} \mathbf{b}^T \mathbf{G}. \quad (5.4)$$

Lihtsamate indeksite puhul saab lähtuda ka viimase maatriksvõrduse elementhaaval esitusest

$$\sigma_I^2 = \sum_{i=1}^m [b_i \times \text{cov}(X_i, A)]. \quad (5.5)$$

Hinnatava aretusväärtuse standardviga avaldub seosena

$$\text{se}(A) = \sigma_I = (\mathbf{b}^T \mathbf{G})^{1/2} = \sqrt{\sum_{i=1}^m [b_i \times \text{cov}(X_i, A)]}.$$

Selektsooniindeksi täpsuse (*accuracy*) all mõistetakse korrelatsiooni indeksi kujul hinnatud ja tegeliku aretusväärtuse vahel, $r_A = r(I, A) = r(\hat{A}, A)$.

Korrelatsioon indeksi ja tegeliku aretusväärtuse vahel esitub vastavalt korrelatsioonikordaja definitsioonile seosena

² Juhul, kui selektsoon baseerub mitmel (n) erineval tunnusel, on loomulik püüda hinnata aretusväärtused neile kõigile samaaegselt, hinnates selleks aretusväärtuste vektorit $\mathbf{A} = (A_1, A_2, \dots, A_n)^T$ (samuti toimitakse juhul, kui korraka hinnatakse aretusväärtusi n loomale); vastavalt muutub ka kovariatsioonimaatriksite dimensioon – aretusväärtuste ja fenotüübi vahelist kovariatsiooni kajastavas maatriksis $\mathbf{G}_{m \times n} = \text{cov}(\mathbf{X}, \mathbf{A})$ vastab üks veerg igale hinnatavale aretusväärtusele (ja rida konkreetsel informatsiooniallikal mõõdetud fenotüübiväärtusele):

$$\text{cov}(\mathbf{X}, \mathbf{A}) = \begin{pmatrix} \text{cov}(X_{11}, A_1) & \text{cov}(X_{12}, A_2) & \dots & \text{cov}(X_{1n}, A_n) \\ \text{cov}(X_{21}, A_1) & \text{cov}(X_{22}, A_2) & \dots & \text{cov}(X_{2n}, A_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_{m1}, A_1) & \text{cov}(X_{m2}, A_2) & \dots & \text{cov}(X_{mn}, A_n) \end{pmatrix}.$$

Kovariatsioonimaatriksi tähistus \mathbf{G} on traditsiooniline selektsooniindeksite teooriale ja ei märgi üldjuhul sama maatriksit kui segamudelite teoorias (Pt 6 ja 7) kasutatav juhulike efektide dispersioonimaatriksi tähistus.

$$r_{IA} = r(I, A) = \frac{\text{cov}(I, A)}{\sqrt{\text{var}(I) \text{var}(A)}}. \quad (5.6)$$

Et

$$\text{cov}(I, A) = \text{cov}[(b_1 X_1 + b_2 X_2 + \dots + b_m X_m), A] = b_1 \text{cov}(X_1, A) + b_2 \text{cov}(X_2, A) + \dots + b_m \text{cov}(X_m, A) = \mathbf{b}^T \mathbf{G} = \sigma_I^2,$$

siis

$$r_{IA} = \sigma_I^2 / \sqrt{\sigma_I^2 \sigma_A^2} = \sigma_I / \sigma_A = \sqrt{\sum_{i=1}^m [b_i \times \text{cov}(X_i, A)]} / \sigma_A. \quad (5.7)$$

Keerulisemate indeksite puhul on sageli mõttekas arvutada nende täpsus otse dispersioonimaatriksite \mathbf{G} ja \mathbf{P} ning populatsiooni geneetiliste parameetrite kaudu, ilma indeksi kordajaid b_i eelnevalt välja arvatamata:

$$r_{IA} = \sqrt{\frac{\mathbf{G}^T \mathbf{P}^{-1} \mathbf{G}}{h^2 \sigma_P^2}}. \quad (5.8)$$

Viimase võrduse tuletamisel on lähtutud sellest, et vastavalt valemeile (5.4) ja (5.3) ning maatriksite transponeerimise omadustele $\sigma_I^2 = \mathbf{b}^T \mathbf{G} = (\mathbf{P}^{-1} \mathbf{G})^T \mathbf{G} = \mathbf{G}^T \mathbf{P}^{-1} \mathbf{G}$, ning päritavuskoeffitsiendi definitsioonist $h^2 = \sigma_A^2 / \sigma_P^2$ tulenevalt $\sigma_A = \sqrt{h^2 \sigma_P^2}$.

Hinnatava aretusväärtuse standardviga avaldub indeksi täpsuse kaudu kujul

$$se(A) = r_{IA} \times \sigma_A. \quad (5.9)$$

5.1.3 Selektiooniindeksi suhteline täpsus *

Erineva informatsiooni selektiooniindeksiks inkorporeerimisel võib sageli osutada vajalikuks teadmine, kuivõrd ühe või teise informatsiooniallika arvesse võtmine või mittevõtmine indeksi täpsust muudab. Kuna igale informatsiooniallikale vastab selektiooniindeksis üks liidetav ja seeläbi ka üks kordaja b_i , on valemite (5.5) ja (5.7) alusel leitavad avaldised indeksi nn suhtelise efektiivsuse tarvis.

Võtame vaatluse alla kaks selektiooniindeksit:

$$I = b_1 X_1 + \dots + b_m X_m$$

ja

$$I_{-i} = b_1 X_1 + \dots + b_{i-1} X_{i-1} + b_{i+1} X_{i+1} + \dots + b_m X_m.$$

Cunningham, E.P. (The relative efficiencies of selection indexes. Acta. Agric. Scand. 19, 45-48, 1969) näitas, et

$$\sigma_I^2 - \sigma_{I_{-i}}^2 = (b_i)^2 / W_{ii}, \quad (5.10)$$

kus W_{ii} on fenotüübil mõõdetud erinevuste dispersioonimaatriksi \mathbf{P} pöördmaatriksi vastav diagonaalielement,

$$\mathbf{P}^{-1} = \mathbf{W} = \begin{pmatrix} W_{11} & \dots & W_{1m} \\ \vdots & \ddots & \vdots \\ W_{m1} & \dots & W_{mm} \end{pmatrix}.$$

Indeksi täpsuse muutus avaldub kujul

$$r_{IA} - r_{I_{-i}A} = \frac{\sqrt{\sigma_I^2} - \sqrt{\sigma_I^2 - (b_i)^2 / W_{ii}}}{\sqrt{\sigma_A^2}}$$

ja **suhteline efektiivsus**, mõõdetuna i . informatsiooniallika X_i osakaaluna kogu indeksi täpsusest, kujul

$$\frac{r_{I_{-i}A}}{r_{IA}} = \sqrt{\frac{\sigma_I^2 - (b_i)^2 / W_{ii}}{\sigma_I^2}}. \quad (5.11)$$

5.2 ARETUSVÄÄRTUSE PROGNOOSIMINE VAID ÜHELE INFORMATSIOONIALLIKALE TUGINEDES

Aretusteoorias defineeritakse indiviidi aretusvärtus enamasti kui tema lõpmatu arvu järglaste keskmise fenotüübiväärtuse \bar{P}_∞ kahekordne erinevus populatsiooni keskmisest \bar{P} :

$$A = 2(\bar{P}_\infty - \bar{P}). \quad (5.12)$$

Kuna reaalses arvutustes ei saa kunagi olla tegu lõpmatu suure järglaste grupiga, kasutatakse täpsemate tulemuste saamiseks kordaja 2 asemel mitmesuguseid järglaste arvu ja uuritava tunnuse geneetilise determineerituse määraga arvestavaid kordajaid, mille kuju leitakse selektiooniindeksi kordajate valemeist (5.2) ja (5.3) lähtuvalt.

Tänu oma lihtsusele leiab valem (5.12) vahel siiski ka praktilist rakendust, saamaks hinnangut indiviidi aretusvärtuse ülemisele piirile, omamata mingit eelinfot uuritava tunnuse geenidest tingituse kohta.

5.2.1 Aretusvärtuse prognoosimine indiviidi enese ühekordselt mõõdetud fenotüübiväärtuse alusel

Kui igal indiviidil on mõõdetud vaid üks selektiooni aluseks oleva tunnuse väärtus, on indiviidi i aretusvärtus selektiooniindeksi definitsioonist (5.1) lähtudes prognoositav valemist

$$\hat{A}_i = b(P_i - \bar{P}), \quad (5.13)$$

kus P_i tähistab fenotüübiväärtust indiviidil i ja \bar{P} populatsiooni keskmist fenotüübiväärtust. Kordaja b määramiseks saame välja kirjutada valemi (5.2):

$$b = \text{cov}(A_i, P_i - \bar{P}) / \text{var}(P_i - \bar{P}) = \text{cov}(A_i, A_i + E_i) / \text{var}(P_i) = \sigma_A^2 / \sigma_P^2 = h^2$$

(sest \bar{P} on konstant, mistõttu $\text{var}(\bar{P}) = 0$, $\text{cov}(A, \bar{P}) = 0$ ja $\text{cov}(A, E) = 0$).

Indeksi täpsus on leitav seosest

$$r_{IA} = \text{cov}(A, P - \bar{P}) / \sqrt{\text{var}(P - \bar{P}) \text{var}(A)} = \sigma_A^2 / (\sigma_P \sigma_A) = \sqrt{h^2} \quad (5.14)$$

ja aretusvärtuse standardviga valemist $se(A) = h^2 \sigma_P$. Seega sõltub see, kui täpselt indiviidi enese fenotüübiväärtus tema aretusvärtust kirjeldab, vaid uuritava tunnuse päritavusest vaadeldavas populatsioonis. Aretusvärtuse hinnangu varieeruvus sõltub lisaks ka veel uuritava tunnuse dispersioonist.

Näide 5.1. Olgu aastavanuse mullika kehamass 320 kg ning kogu karja keskmine vastav näitaja 250 kg. Arvutame mullika aretusvärtuse ja saadud hinnangu täpsuse eeldusel, et aastase kehamassi päritavus on 0,45.

Vastavalt valemeile (5.13) ja (5.14) saame:

$$\hat{A} = 0,45 \times (320 - 250) = 31,5 \text{ kg}, \quad r_{IA} = \sqrt{0,45} = 0,67.$$

5.2.2 Aretusvärtuse prognoosimine indiviidi enese korduvalt mõõdetud fenotüübiväärtuste alusel

Eeldame, et indiviidil on selektiooni aluseks oleval tunnusel sooritatud n mõõtmist ning kõigi mõõtmiste dispersioon on sama nagu ka kõigi mõõtmiste vahelised keskkonnamõjudest tingitud korrelatsioonid (kõik sama looma mõõtmiste vahelised geneetilised korrelatsioonid võrduvad ühega, sest tegu on ju ühe ja sama tunnusega, mis on mõjutatud samade geenide poolt).

Tähistame indiviidi i mõõtmiste keskmise \bar{P}_i . Vaatlusaluse looma aretusvärtus on siis avaldatav seosenä

$$\hat{A}_i = b(\bar{P}_i - \bar{P}), \quad (5.15)$$

kus

$$b = \text{cov}(A_i, \bar{P}_i - \bar{P}) / \text{var}(\bar{P}_i - \bar{P}) = \text{cov}(A_i, \bar{P}_i) / \text{var}(\bar{P}_i).$$

Et vastavalt valemile (3.11) avaldub indiviidil i sooritatud j . mõõtmise tulemus P_{ij} kujul $P_{ij} = \bar{P} + G_i + E_{pi} + E_{ij}$ (genotüübiefekt G_i ja püsiv keskkonnaefekt E_{pi} on kõigil mõõtmistel samad ja

ei sõltu seega konkreetse mõõtmise järjekorranumbrist j), siis avaldub n mõõtmise keskmine seosena $\bar{P}_i = \bar{P} + G_i + E_{pi} + \frac{1}{n} \sum_{j=1}^n E_{ij}$, kus E_{ij} on mõõtmisele j vastav juhuslik keskkonnamõju. Seega

$$\text{cov}(A_i, \bar{P}_i) = \text{cov}(A_i, G_i + E_{pi} + \sum_n E_{ij}/n) = \text{cov}(A_i, G_i) = \sigma_A^2$$

ja, eeldades genotüübi ning püsiva ja juhuslikku keskkonnamõju sõltumatust,

$$\text{var}(\bar{P}_i) = \text{var}(G_i) + \text{var}(E_{pi}) + \text{var}\left(\sum_{j=1}^n E_{ij}/n\right).$$

Arvestades, et kõik juhuslikud vead E_{ij} on eeldatavalt sama dispersiooniga, avaldub viimane liidetav vastavalt dispersiooni omadustele kujul

$$\text{var}\left(\frac{1}{n} \sum_{j=1}^n E_{ij}\right) = \frac{1}{n^2} \sum_{j=1}^n [\text{var}(E_{ij})] = \frac{1}{n^2} n \text{var}(E) = \text{var}(E)/n.$$

Et vastavalt korduvuse definitsioonile (3.12) $R = [\text{var}(G) + \text{var}(E_p)]/\text{var}(P)$ ja sellest järelduvalt

$$1 - R = \frac{\text{var}(P)}{\text{var}(P)} - \frac{\text{var}(G) + \text{var}(E_p)}{\text{var}(P)} = \frac{\text{var}(E)}{\text{var}(P)},$$

siis

$$\text{var}(\bar{P}_i) = \frac{\text{var}(P) [\text{var}(G) + \text{var}(E_p)]}{\text{var}(P)} + \frac{\text{var}(P) \text{var}(E)}{n \text{var}(P)} = \sigma_P^2 R + \frac{\sigma_P^2 (1-R)}{n} = \left[R + \frac{(1-R)}{n} \right] \sigma_P^2.$$

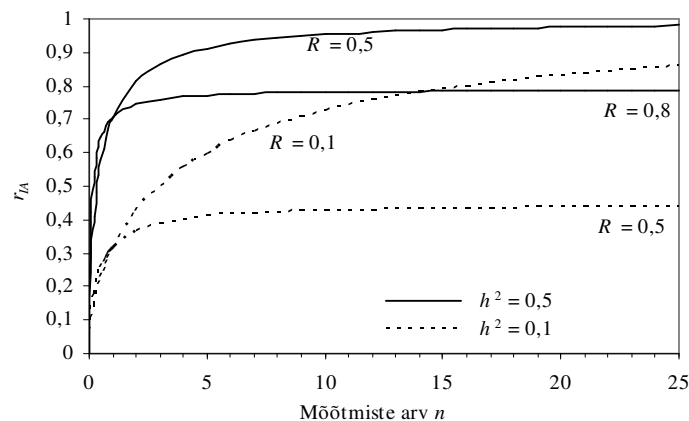
Seega avaldub selektsiooniindeksi (5.15) kordaja b kokkuvõtteks seosest

$$b = \frac{\sigma_A^2}{[R + (1-R)/n] \sigma_P^2} = \frac{h^2}{[nR + 1 - R]/n} = \frac{nh^2}{[1 + (n-1)R]}.$$

Indeksi täpsus on tänu seosele (5.7) leitav valemist

$$r_{IA} = \sqrt{b \times \text{cov}(A_i, \bar{P}_i)} / \sigma_A = \sqrt{b \times \sigma_A^2} / \sigma_A = \sqrt{b} = \sqrt{\frac{nh^2}{[1 + (n-1)R]}}.$$

Joonisel 5.1 on kujutatud indiviidi aretusväärtuse hinnangu täpsus sõltuvalt mõõtmiste arvust ning päritavuskoefitsiendi ja korduvuse väärtustest. Mida suurem on päritavuskoefitsiendi väärtus ja mida rohkem on sooritatud mõõtmisi, seda täpsem on aretusväärtuse hinnang. Korduvuse kõrge väärtus pärsib lisamõõtmiste positiivset mõju aretusväärtuse hinnangu täpsusele – see on ka loomulik, sest kõrge korduvuse korral on kõigi samal indiviidil sooritatud mõõtmiste tulemused üsna sarnased ega sisalda märkimisväärselt lisainformatsiooni, mida saaks ära kasutada aretusväärtuse hindamiseks. Lisamõõtmiste positiivne efekt aretusväärtuse hinnangu täpsusele ilmneb eelkõige madala päritavuse ja korduvuse korral, kõrge päritavuse ja eelkõige kõrge korduvuse korral ei ole lisamõõtmistel erilist mõtet.



Joonis 5.1. Indiviidi aretusväärtuse hinnangu täpsus sõltuvalt mõõtmiste arvust ning päritavuskoefitsiendi ja korduvuse väärtustest.

Näide 6.2. Eeldame, et lehm Roosi kolme esimese laktatsioonikuu keskmine piimatoodang oli 7000 kg. Leiame Roosi piimatoodangu aretusväärtuse eeldusel, et karja keskmine sama perioodi piimatoodang oli 6000 kg, esimese 3 kuu piimatoodangu päritavus on 0,3 ja korrelatsioon 3 kuu toodangute vahel (korduvus) on 0,5.

Vastavalt valemile (5.15) saame:

$$\hat{A} = b(7000 - 6000),$$

kus

$$b = 3 \times (0,3) / [1 + (3 - 1) \times 0,5] = 0,45,$$

millest

$$\hat{A} = 0,45 \times (7000 - 6000) = 450 \text{ kg}.$$

Aretusväärtuse hinnangu täpsuseks saame $r_{IA} = \sqrt{0,45} = 0,67$.

5.2.3 Aretusväärtuse prognoosimine järglaste fenotüübiväärtuste alusel

Kui uuritava tunnuse väärtused saavad olla mõõdetud üksnes emastel indiviididel, toimub isade aretusväärtuste hindamine enamasti nende järglaste keskmiste väärtuste alusel.

Tähistame vaadeldava isa järglaste keskmist väärtust \bar{P}_s ja eeldame, et järglased on omavahel sugulased vaid isa kaudu. Vaatlusaluse isa aretusväärtus A_s on siis hinnatav seosest

$$\hat{A}_s = b(\bar{P}_s - \bar{P}) \quad (5.16)$$

kus $b = \text{cov}(A_s, \bar{P}_s) / \text{var}(\bar{P}_s)$.

Et $\text{cov}(A_s, \bar{P}_s) = \text{cov}[A_s, \frac{1}{2} A_s + \frac{1}{n} \sum_{i=1}^n (\frac{1}{2} A_{di}) + \frac{1}{n} \sum_{i=1}^n E_i]$, kus A_s ja A_{di} märgivad vastavalt isa ja ema aretusväärtusi ning n tähistab järglaste arvu vaatlusalusel isal, ja et meile pakub huvi üksnes isa poolt järglasele edasi kandunud geneetiline materjal (järglastele mõjuvad emapoolsed aditiivgeneetilised efektid nagu ka kõik mitteaditiivsed geneetilised efektid loetakse kuuluvaks juhuslike keskkonnamõjude hulka, mis uuritava isa aretusväärtusega ei korreleeru), siis

$$\text{cov}(A_s, \bar{P}_s) = \text{cov}(A_s, \frac{1}{2} A_s) = \frac{1}{2} \sigma_A^2.$$

Arvestades, et sama isa aga erinevate emade järglaste aretusväärtuste sarnasus on tingitud just isalt pärandunud geenidest, esitatakse ka järglaste keskmise fenotüübiväärtuse dispersioon $\text{var}(\bar{P}_s)$ isalt pärandunud geenidest tingitud varieeruvuse kaudu:

$$\begin{aligned} \text{var}(\bar{P}_s) &= \text{var}\left[\bar{P} + \frac{1}{2} A_s + \frac{1}{n} \sum_{i=1}^n (\frac{1}{2} A_{di} + E_i)\right] = \text{var}(\frac{1}{2} A_s) + \text{var}\left[\frac{1}{n} \sum_{i=1}^n (\frac{1}{2} A_{di} + E_i)\right] \\ &= \frac{1}{4} \text{var}(A_s) + \frac{1}{n^2} \times \sum_{i=1}^n [\text{var}(\frac{1}{2} A_{di} + E_i)] = \frac{1}{4} \sigma_A^2 + \frac{1}{n} (\sigma_P^2 - \frac{1}{4} \sigma_A^2). \end{aligned}$$

Siin $\frac{1}{4} \sigma_A^2$ on isa aditiivgeneetilisest mõjust tingitud dispersioon ja $\sigma_P^2 - \frac{1}{4} \sigma_A^2$ isalt pärandunud geenide summaarse mõjuga mitte kirjeldatav osa fenotüübilisest varieeruvusest σ_P^2 (viimane hõlmab nii emalt pärandunud geenide mõjust kui ka keskkonnatingimustest tingitud varieeruvust fenotüübiväärtustes). Et päritavuskoeffitsient $h^2 = \sigma_A^2 / \sigma_P^2$, siis järelikult

$$\frac{1}{4} \sigma_A^2 / \sigma_P^2 = \frac{1}{4} h^2 \quad \text{ja} \quad (\sigma_P^2 - \frac{1}{4} \sigma_A^2) / \sigma_P^2 = 1 - \frac{1}{4} h^2.$$

Viimastest võrdustest tulenevalt

$$\text{var}(\bar{P}_s) = \frac{1}{4} h^2 \sigma_P^2 + \frac{1}{n} (1 - \frac{1}{4} h^2) \sigma_P^2, \quad (5.17)$$

millest kordaja indeksis (5.16) saab kuju

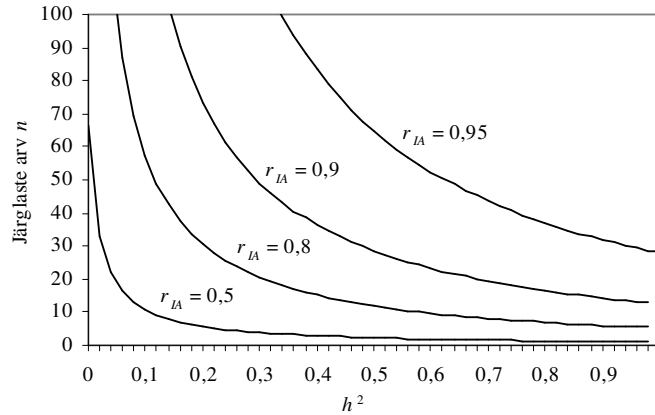
$$b = \frac{\frac{1}{2} \sigma_A^2}{\left[\frac{1}{4} h^2 + \frac{1}{n} (1 - \frac{1}{4} h^2)\right] \sigma_P^2} = \frac{\frac{1}{2} n h^2}{\left[\frac{1}{4} n h^2 + (1 - \frac{1}{4} h^2)\right]} = \frac{\frac{1}{2} n h^2}{1 + \frac{1}{4} h^2 (n - 1)}. \quad (5.18)$$

Kui nüüd eeldada, et järglaste arv $n \rightarrow \infty$, siis $1 + \frac{1}{4} h^2 (n - 1) \rightarrow \frac{1}{4} n$ ja $\frac{1}{2} n h^2 \rightarrow \frac{1}{2} n$, mistõttu selektiooniindeksi (5.16) valemiga (5.18) defineeritud kordaja $b = [\frac{1}{2} n h^2] / [1 + \frac{1}{4} h^2 (n - 1)] \rightarrow \frac{1}{2} n / \frac{1}{4} n = 2$, andes tulemuseks aretusväärtuse definitsiooni (5.12).

Isa aretusväärtuse hinnangu täpsus avaldub vastavalt valemile (5.7) kujul

$$r_{IA} = \sqrt{\frac{\frac{1}{2} \sigma_A^2 \times \frac{\frac{1}{2} n h^2}{1 + \frac{1}{4} h^2 (n - 1)}}{\sigma_A^2}} = \sqrt{\frac{\frac{1}{4} n h^2}{1 + \frac{1}{4} h^2 (n - 1)}}.$$

On loomulik eeldada, et isa aretusväärtuse hinnang on seda täpsem, mida enam on isal järglasi. Joonisel 5.2 on kujutatud erinevate täpsuste saavutamiseks vajalik järglaste arv sõltuvalt päritavuskoeffitsiendi väärtusest. Imneb, et mida suurem on päritavuskoeffitsiendi väärtus, seda vähem on vaja järglasi isa (või ema) aretusväärtuse täpseks hindamiseks.



Joonis 5.2. Isa aretusväärtuse hinnangu täpsus sõltuvalt järglaste arvust ja päritavuskoeffitsiendi väärtusest.

5.2.4 Aretusväärtuse prognoosimine poolõvede fenotüübiväärtuste alusel

Vaadeldava isa iga tulevase tütre potentsiaalne suutlikkus (aretusväärtus) on prognoositav praeguste tütarde keskmise suutlikkuse \bar{P}_s alusel seosest

$$\hat{A}_{daugh} = b(\bar{P}_s - \bar{P}), \quad (5.19)$$

kus $b = \text{cov}(A_{daugh}, \bar{P}_s) / \text{var}(\bar{P}_s)$.

Kovariatsioon tulevase tütre aretusväärtuse ja praeguste järglaste keskmise fenotüübiväärtuse vahel avaldub seosena $\text{cov}(A_{daugh}, \bar{P}_s) = \text{cov}(\frac{1}{2}A_s + \frac{1}{2}A_{d^*}, \frac{1}{2}A_s + \frac{1}{2}A_d + \sum_n E/n)$, kus d^* tähistab tulevase tütre ema, kes eeldatavalt ei ole suguluses praeguste tütarde emadega (d). Et ka isad ja emad ei ole eeldatavalt omavahel sugulased ja puudub korrelatsioon keskkonna ja genotüübi vahel, jääb tulevase tütre aretusväärtuse ja praeguste tütarde keskmise fenotüübiväärtuse kovariatsioonis alles vaid üks nullist erinev liidetav:

$$\begin{aligned} \text{cov}(A_{daugh}, \bar{P}_s) &= \text{cov}\left\{\left(\frac{1}{2}A_s + \frac{1}{2}A_{d^*}\right), \left[\bar{P} + \frac{1}{2}A_s + \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{2}A_{d_i}\right) + \frac{1}{n}\sum_{i=1}^n E_i\right]\right\} \\ &= \text{cov}\left(\frac{1}{2}A_s, \frac{1}{2}A_s\right) = \frac{1}{4}\text{cov}(A_s, A_s) = \frac{1}{4}\sigma_A^2. \end{aligned}$$

Viimase tulemuse ja võrduse (5.17) alusel saab kordaja b indeksis (5.19) kuju

$$b = \frac{\frac{1}{4}\sigma_A^2}{\left[\frac{1}{4}h^2 + \left(1 - \frac{1}{4}h^2\right)/n\right]\sigma_P^2} = \frac{\frac{1}{4}nh^2}{1 + \frac{1}{4}h^2(n-1)}. \quad (5.20)$$

Ainuke erinevus avaldiste (5.18) ja (5.20) vahel on, et viimase lugejas on kordaja 2 kordaja väiksem, mis tähendab, et tulevase tütre aretusväärtus on pool isa aretusväärtusest. Tulevase tütre fenotüübiväärtus on prognoositav seosest

$$\hat{P}_{daugh} = \mu + \hat{A}_{daugh} = \mu + 0,5\hat{A}_s, \quad (5.21)$$

kus μ on karja/populatsiooni keskmine fenotüübiväärtus ning \hat{A}_{daugh} leitakse valemist (5.19), kus kordaja b on määratud avaldisega (5.20). Tulevase tütre aretusväärtuse hinnangu \hat{A}_{daugh} täpsus on

$$r_{A_{daugh}, \hat{A}_{daugh}} = \sqrt{\frac{\frac{1}{4}nh^2}{1 + \frac{1}{4}h^2(n-1)}},$$

mis on pool isa aretusväärtuse hinnangu täpsusest.

Näide 5.3. Olgu pulli Elroi 25 tütre keskmine 1. laktatsiooni rasvatoodang 200 kg. Vastav karja keskmine näitaja on 230 kg ja rasvatoodangu päritavus 0,3. Leiame pulli rasvatoodangu aretusväärtuse ja tema tulevaste tütarde oletatava rasvatoodangu samas karjas.

Vastavalt valemile (5.16) saame:

$$\hat{A}_{Elroi} = b(200 - 230),$$

kus

$$b = (\frac{1}{4} \times 25 \times 0,3) / [1 + \frac{1}{4} \times 0,3 \times (25 - 1)] = 1,34,$$

millest

$$\hat{A}_{Elroi} = 1,34(200 - 230) = -40,2 \text{ kg.}$$

Leitud hinnangu korrelatsioon tegeliku aretusväärtusega on $r_{IA} = \sqrt{\frac{1}{2} \times 1,34} \approx 0,82$.

Elroi tulevaste tütarde oletatav 1. laktatsiooni rasvatoodang samas karjas on vastavalt valemile (5.21) $\hat{P}_{daugh} = \mu + 0,5\hat{A}_s = 230 + 0,5(-40,2) = 209,9 \text{ kg}$,

tütarde aretusväärtuse hinnang on $\hat{A}_{daugh} = \frac{1}{2} \times (-40,2) = -20,1 \text{ kg}$,

täpsusega $r_{\hat{A}_{daugh}, \hat{A}_{daugh}} \approx \frac{1}{2} \times 0,82 = 0,41$.

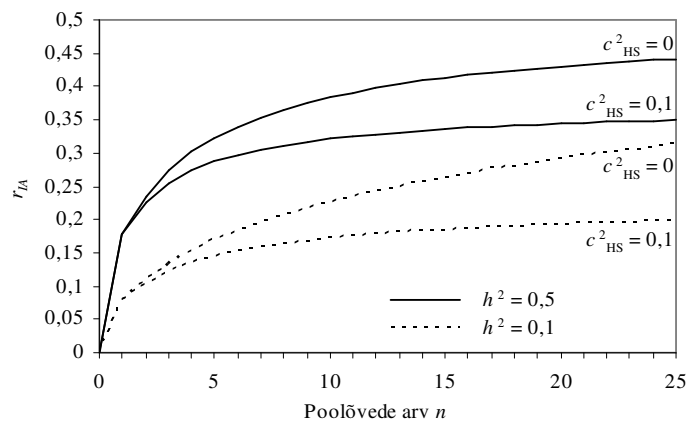
Juhul, kui vaatlusaluse indiviidi poolõvedele omased mittegeneetilised mõjud on mingil määral sarnased (näiteks on tegu mingi noore testpulli ühel ja samal aastal ning teatud valitud farmides sündinud järglastega), siis sõltub keskmise fenotüübiväärtuse dispersioon $\text{var}(\bar{P}_s)$ lisaks aditiivgeneetilise mõju osakaalu näitavale päritavuskoeffitsiendile h^2 ka poolõvedele ühise mitteaditiivgeneetilise mõju osast fenotüübilises varieeruvuses c_{HS}^2 (fenotüübidiispersioon eeldatakse esituvat komponentidena $\sigma_P^2 = \sigma_A^2 + \sigma_{HS}^2 + \sigma_E^2$, millest $h^2 = \sigma_A^2 / \sigma_P^2$ ja $c_{HS}^2 = \sigma_{HS}^2 / \sigma_P^2$).

Selektiooniindeksi (5.19) valemiga (5.20) avalduv kordaja teiseneb siis kujule

$$b = \frac{\frac{1}{4}nh^2}{1 + (n-1)(\frac{1}{4}h^2 + c_{HS}^2)}.$$

Kui poolõvedele ühine mitteaditiivgeneetilise mõju puudub, siis $c_{HS}^2 = 0$ ja kordaja b avaldub valemiga (5.20).

Joonisel 5.3 on kujutatud poolõvede fenotüübiväärtuste alusel leitud aretusväärtuse hinnangu täpsus sõltuvalt poolõvede arvust, päritavuskoeffitsiendi väärtusest ja poolõvedele ühise mitteaditiivgeneetilise mõju osast. Silma hakkab, et kuigi poolõvede arvu kasvades suureneb ka aretusväärtuse hinnangu täpsus, on poolõvede arvu efekt suurem kõrgema päritavuse ja madalama poolõvedele ühise mitteaditiivgeneetilise mõju osa korral. Seejuures suureneb aretusväärtuse hinnangu täpsus kiirelt poolõvede arvu suurenemisel 3-4-ni, edasine täpsuse suurenemine on juba väike ja seda eriti juhul, kui poolõved on omavahel sarnased lisaks samalt vanemalt päritud geenidele ka ühiste keskkonnamõjude tõttu.



Joonis 5.3. Poolõvede fenotüübiväärtuste alusel leitud aretusväärtuse hinnangu täpsus sõltuvalt poolõvede arvust, päritavuskoeffitsiendi väärtusest ja poolõvedele ühise mitteaditiivgeneetilise mõju osast.

5.2.5 Kokkuvõtlikult aretusväärtuse prognoosimisest sugulaste fenotüübiväärtuste alusel

Mingi generatsiooni t kuuluva indiviidi aretusväärtuse hindamisel kasutatavad sugulaste fenotüübiväärtused võib laias laastus jagada kolme gruppi:

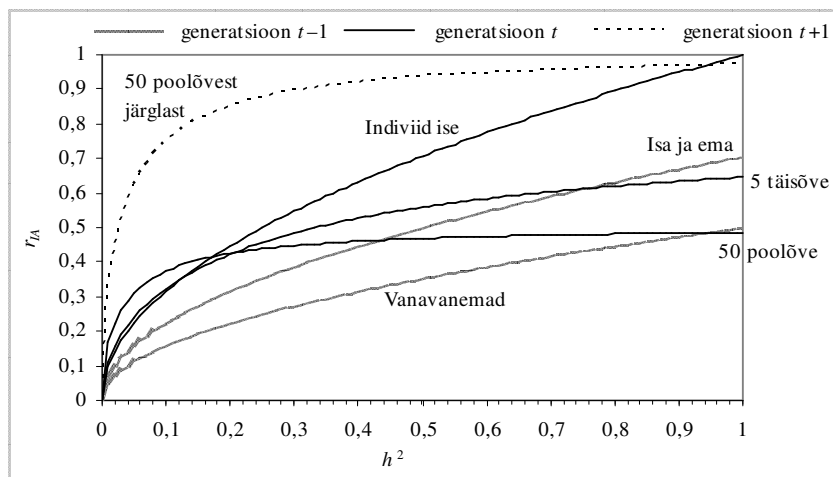
- generatsiooni $t-1$ (ja vanematesse generatsioonidesse) kuuluvate eellaste (vanemate, vanavanemate jne) fenotüübiandmed;
- generatsiooni t kuuluvate indiviidi enese või tema pool- või täisõvede fenotüübiandmed;
- generatsiooni $t+1$ (ja edasistesse generatsioonidesse) kuuluvate järglaste fenotüübiandmed.

Tabelisse 5.1 on koondatud selektsiooniindeksi kordaja b ja täpsuse r_{IA} avaldised tüüpilisemate informatsiooniallikate korral (eeldusel, et aretusväärtuse hinnang põhineb üksnes antud infol).

Joonisel 5.4 on aga esitatud aretusväärtuse hinnangu täpsus sõltuvalt päritavuskoefitsiendi väärtusest ja erinevatel ajajärgudel kättesaadavast informatsioonist.

Tabel 5.1. Selektiiooniindeksi kordaja b ja täpsuse r_{IA} avaldised tüüpilisemate informatsiooniallikate korral (h^2 on päritavus, R korduvus, c_{FS}^2 ja c_{HS}^2 vastavalt täis- ja poolõvedele ühiste mitteadiitiivgeneetiliste mõjude osa kogu fenotüübilisest varieeruvusest).

Informatsiooniallikas (kasutatav fenotüübiväärtus)	Selektiiooniindeksi kordaja (b)	Selektiiooniindeksi täpsus (r_{IA})
Vanavanemad (4 väärtuse keskmine)	h^2	$\sqrt{\frac{1}{4}h^2}$
Vanemad (2 väärtuse keskmine)	h^2	$\sqrt{\frac{1}{2}h^2}$
Indiviid ise (1 väärtus)	h^2	$\sqrt{h^2}$
Indiviid ise (n korduva mõõtmise keskmine)	$\frac{nh^2}{1+(n-1)R}$	$\sqrt{\frac{nh^2}{1+(n-1)R}}$
Täisõved (n mõõtmise keskmine)	$\frac{\frac{1}{2}nh^2}{1+(n-1)(\frac{1}{2}h^2+c_{FS}^2)}$	$\sqrt{\frac{\frac{1}{4}nh^2}{1+(n-1)(\frac{1}{2}h^2+c_{FS}^2)}}$
Poolõved (n mõõtmise keskmine)	$\frac{\frac{1}{4}nh^2}{1+(n-1)(\frac{1}{4}h^2+c_{HS}^2)}$	$\sqrt{\frac{\frac{1}{16}nh^2}{1+(n-1)(\frac{1}{4}h^2+c_{HS}^2)}}$
Järglased (n mõõtmise keskmine)	$\frac{\frac{1}{2}nh^2}{1+\frac{1}{4}h^2(n-1)}$	$\sqrt{\frac{\frac{1}{4}nh^2}{1+\frac{1}{4}h^2(n-1)}}$



Joonis 5.4. Aretusväärtuse hinnangu täpsus sõltuvalt päritavuskoefitsiendi väärtusest ja erinevatel ajajärgudel kättesaadavast informatsioonist (püsivad keskkonnaefektid on loetud võrdseks nulliga).

5.2.6 Aretusväärtuse prognoosimine ühele tunnusele teise kaudu

Aretusväärtused ühele tunnusele on prognoositavad teise tunnuse kaudu, kui nende tunnuste vaheline geneetiline korrelatsioon erineb nullist. Olgu P_{Z_i} tunnuse Z väärtus mõõdetuna indiviidil i ja \bar{P}_Z selle tunnuse keskmine fenotüübiväärtus populatsioonis. Indiviidi i aretusväärtus tunnusele Y on siis leitav selektiooniindeksist

$$I = \hat{A}_{Y_i} = b(P_{Z_i} - \bar{P}_Z), \quad (5.22)$$

kus $b = \text{cov}(A_Y, P_Z) / \text{var}(P_Z)$.

Kovariatsioon viimases võrduses on vastavalt lineaarsele geneetilisele mudelile (4.3) ja geneetilise korrelatsiooni definitsioonile (4.12) esitatav seosena

$$\text{cov}(A_Y, P_Z) = \text{cov}(A_Y, \bar{P}_Z + A_Z + E_Z) = \text{cov}(A_Y, A_Z) = r_{G_{YZ}} \sqrt{\sigma_{A_Y}^2 \sigma_{A_Z}^2},$$

kus $r_{G_{YZ}}$, $\sigma_{A_Y}^2$ ja $\sigma_{A_Z}^2$ tähistavad vastavalt geneetilist korrelatsiooni tunnuste Y ja Z vahel ning tunnuste Y ja Z aditiivdispersioone.

Et aditiivdispersioon on päritavuse abil esitatav seosena $\sigma_A^2 = h^2 \sigma_P^2$, avaldub selektiooniindeksi kordaja b kujul

$$b = r_{G_{YZ}} \sigma_{A_Y} \sigma_{A_Z} / \sigma_{P_Z}^2 = r_{G_{YZ}} \sigma_{P_Y} \sigma_{P_Z} h_Y h_Z / \sigma_{P_Z}^2 = r_{G_{YZ}} \sigma_{P_Y} h_Y h_Z / \sigma_{P_Z}$$

(siin $h = \sqrt{h^2}$ ja $\sigma = \sqrt{\sigma^2}$).

Hinnatud aretusväärtuse täpsus avaldub kujul

$$r_{I_{A_Y}} = \sqrt{\frac{r_{G_{YZ}}^2 \sigma_{A_Y}^2 \sigma_{A_Z}^2}{\sigma_{P_Z}^2 \sigma_{A_Y}^2}} = |r_{G_{YZ}} h_Z|.$$

Näide 5.4. Võtame vaatluse alla 2 tunnust – veiste keskmise ööpäevase massi-iibe (P_1) ja söödatarbe 1 kg juurdekasvu saavutamiseks (söödakasutuse efektiivsuse, P_2) mõõdetuna erinevusena populatsiooni keskmisest. Enamasti on looma söödakasutuse efektiivsuse reaalne mõõtmine komplitseeritud või suisa võimatu ja indiviidide üksnes selle tunnuse alusel järjestamine seetõttu keeruline. Teades aga, et looma ööpäevane juurdekasv ja söödakasutusvõime on tugevalt korreleeritud, võime eeldada muutusi loomade söödakasutusvõimes ka peale üksnes ööpäevasel juurdekasvul baseeruvat seleksiooni.

Omades eelnevatest uuringutest andmeid mõlema tunnuse fenotüübdispersioonide ja päritavuskoeffitsientide ning tunnustevahelise geneetilise korrelatsiooni kohta: $\sigma_{P_1}^2 = 36$, $\sigma_{P_2}^2 = 64$, $h_{P_1}^2 = 0,44$, $h_{P_2}^2 = 0,39$ ja $r_{G_{12}} = -0,65$, saame igale indiviidile hinnata valemist (5.22) tema söödakasutusvõime aretusväärtuse.

Indeksi (5.22) kordaja saame seosest

$$b = r_{G_{12}} \sigma_{P_2} h_{P_2} h_{P_1} / \sigma_{P_1} = -0,65 \times 8 \times \sqrt{0,39 \times 0,44} / 6 = -0,36$$

ja indeks ise on kujul

$$I = \hat{A}_2 = -0,36 P_1.$$

Asendades viimases valemis kordaja P_1 konkreetse looma ööpäevase juurdekasvu erinevusega populatsiooni keskmisest, saame teada looma söödakasutuse efektiivsuse aretusväärtuse.

Negatiivne selektiooniindeksi kaaluparameeter on indikaator sellest, et selekteerides loomi suurema ööpäevase juurdekasvu alusel selekteerime me tegelikult ka parema söödakasutusvõimega loomi (neid, kelle puhul söödakulu on väiksem – mida suurem on P_1 , seda väiksem on \hat{A}_2).

Hinnatud aretusväärtuse täpsus on $r_{I_{A_2}} = |r_{G_{12}} h_{P_1}| = |-0,65 \times \sqrt{0,44}| = |-0,43| = 0,43$.

5.3 NÄITEID MITMELE INFORMATSIOONIALLIKALE TUGINEVATEST SELEKTSIOONIINDEKSITEST

Mitmele informatsiooniallikale tuginevate aretusväärtuse hinnangute puhul leitakse selektsiooniindeksi parameetrid maatriksvõrdusest (5.3). Järgnevalt mõned näited vajalike kovariatsioonimaatriksite konstrueerimisest ja nende alusel indeksite väärtuste arutamiseks.

5.3.1 Seleksiooniindeks korreleeruvate tunnuste korral

Juhul, kui selektsioon baseerub ühel tunnusel, mõõdetud on aga mitut tunnust, mis selektsiooni aluseks olevaga tugevalt korreleeruvad ja võimaldavad seeläbi täpsustada aretusväärtuse hinnangut, on mõttekas kaasata selektsiooniindeksisse kõik tunnused, mille kohta infot omatakse.

Olgu meil mõõdetud fenotüüp kahe tunnuse osas, P_1 ja P_2 , ning baseerugu selektsioon teisel tunnusel. Eesmärk on konstrueerida selektsiooniindeks prognoosimaks aretusväärtusi tunnusele P_2 arvestades mõlema tunnuse mõõtmisi:

$$I = \hat{A}_2 = b_1 P_1 + b_2 P_2.$$

Kaaluparameetrite b_1 ja b_2 arutamiseks peame välja kirjutama maatriksvõrduse (5.3):

$$\begin{pmatrix} \sigma_{P_1}^2 & \sigma_{P_1 P_2} \\ \sigma_{P_1 P_2} & \sigma_{P_2}^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sigma_{A_2 P_1} \\ \sigma_{A_2 P_2} \end{pmatrix}.$$

Võrduse paremal poolel asuva vektori elemendid avalduvad lineaarse geneetilise mudeli, päritavuskoefitsiendi ja geneetilise korrelatsioonikordaja definitsioonidest lähtuvalt kujul

$$\sigma_{A_2 P_1} = \text{cov}(A_2, \bar{P}_1 + A_1 + E_1) = \text{cov}(A_2, A_1) = \sigma_{A_2 A_1} = r_{G_{12}} \sigma_{A_1} \sigma_{A_2} = r_{G_{12}} \sigma_{P_1} \sigma_{P_2} h_1 h_2$$

ja

$$\sigma_{A_2 P_2} = \text{cov}(A_2, \bar{P}_2 + A_2 + E_2) = \text{cov}(A_2, A_2) = \sigma_{A_2}^2 = h_2^2 \sigma_{P_2}^2.$$

Võrduse vasakpoolses maatriksis paiknevad fenotüübilised kovariatsioonid saab avaldada fenotüübiliste korrelatsioonide kaudu: $\sigma_{P_1 P_2} = r_{P_{12}} \sigma_{P_1} \sigma_{P_2}$. Kokkuvõttes saame indeksi kaaluparameetrid leida maatriksvõrdusest $\mathbf{b} = \text{var}(\mathbf{X})^{-1} \text{cov}(\mathbf{X}, A) = \mathbf{P}^{-1} \mathbf{G}$:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sigma_{P_1}^2 & r_{P_{12}} \sigma_{P_1} \sigma_{P_2} \\ r_{P_{12}} \sigma_{P_1} \sigma_{P_2} & \sigma_{P_2}^2 \end{pmatrix}^{-1} \begin{pmatrix} r_{G_{12}} \sigma_{P_1} \sigma_{P_2} h_1 h_2 \\ h_2^2 \sigma_{P_2}^2 \end{pmatrix}. \quad (5.23)$$

Näide 5.5. Jätkame näidet 5.5 veiste ööpäevasest juurdekasvust (P_1) ja söödakasutuse efektiivsusest (P_2).

Oletame nüüd, et mõlema tunnuse väärtused on mõõdetud ning selektsiooni aluseks olevaks tunnuseks, millele tahame hinnata aretusväärtusi, on endiselt söödakasutusvõime. Olgu lisaks juba toodud parameetritele ($\sigma_{P_1}^2 = 36$, $\sigma_{P_2}^2 = 64$, $h_{P_1}^2 = 0,44$, $h_{P_2}^2 = 0,39$ ja $r_{G_{12}} = -0,65$) teada ka fenotüübiline korrelatsioon tunnuste vahel, $r_{P_{12}} = -0,83$.

Asendades dispersioonid ja kovariatsioonid valemis (5.23) nende arväärtustega, saame indeksi kaaluparameetriteks

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 36 & -40 \\ -40 & 64 \end{pmatrix}^{-1} \begin{pmatrix} -13 \\ 25 \end{pmatrix} = \begin{pmatrix} 0,24 \\ 0,54 \end{pmatrix},$$

mille alusel selektsiooniindeks on

$$\hat{A}_2 = 0,24 P_1 + 0,54 P_2.$$

Sellisel kujul leitud indeksis vastab nii suuremale juurdekasvule kui ka suuremale söödakulule suurem indeksi väärtus (vaikimisi konstrueeritakse selektsiooniindeksid alati nn positiivses suunas – mida suurem on indiviidil mõõdetud tunnuse väärtus võrreldes populatsiooni keskmisega, seda parem). Et parimateks loetakse ikkagi vähima söödakuluga loomi, tuleks antud juhul selekteerida välja väiksema indeksi väärtusega (söödakulu aretusväärtusega) loomad.

Hinnatud aretusväärtuse täpsus on valemist (5.8) lähtuvalt

$$r_{A_2} = \frac{1}{\sqrt{0,39 \times 64}} \left[(-13 \ 25) \begin{pmatrix} 36 & -40 \\ -40 & 64 \end{pmatrix}^{-1} \begin{pmatrix} -13 \\ 25 \end{pmatrix} \right]^{1/2} = 0,64525.$$

5.3.2 Aretusväärtuse prognoosimine indiviidi enese ja tema vanemate ühekordselt mõõdetud fenotüübiväärtuste alusel

Olgu eesmärgiks hinnata indiviidi aretusväärtust temal enesel ja tema emal ning isal sooritatud mõõtmiste alusel. Tähistades indiviidi enese ja tema ema ning isa fenotüübiväärtuste erinevused populatsiooni keskmisest kui P_o , P_d ja P_s , saame aretusväärtuse hindamiseks kasutatava selektiooniindeksi kujul

$$I = \hat{A}_o = b_1 P_o + b_2 P_d + b_3 P_s. \quad (5.24)$$

Kaaluparameetrite b_1 , b_2 ja b_3 arvutamiseks saame maatriksvõrduse $\underbrace{\text{var}(\mathbf{X})}_{\mathbf{P}} \times \mathbf{b} = \underbrace{\text{cov}(\mathbf{X}, A)}_{\mathbf{G}}$:

$$\begin{pmatrix} \sigma_{P_o}^2 & \sigma_{P_o P_d} & \sigma_{P_o P_s} \\ \sigma_{P_o P_d} & \sigma_{P_d}^2 & \sigma_{P_d P_s} \\ \sigma_{P_o P_s} & \sigma_{P_d P_s} & \sigma_{P_s}^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} \sigma_{A_o P_o} \\ \sigma_{A_o P_d} \\ \sigma_{A_o P_s} \end{pmatrix}.$$

Eeldades, et kõik fenotüübiväärtused on leitud ühes ja samas populatsioonis, on kõik fenotüübidispersioonid võrdsed:

$$\sigma_{P_o}^2 = \sigma_{P_d}^2 = \sigma_{P_s}^2 = \sigma_P^2.$$

Oletades, et kogu vanema ja järglase vaheline kovariatsioon on aditiivgeneetiline ning et isa ja ema pole omavahel sugulased, ja teades, et pooled oma geenidest on järglane pärinud ühelt vanemalt ja pooled teiselt vanemalt, avalduvad fenotüübilised kovariatsioonid kujul

$$\sigma_{P_o P_d} = \sigma_{P_o P_s} = \frac{1}{2} \sigma_A^2 = \frac{1}{2} h^2 \sigma_P^2 \quad \text{ja} \quad \sigma_{P_d P_s} = 0.$$

Analoogsete arutelude tulemusena saame, et

$$\sigma_{A_o P_o} = \sigma_A^2 = h^2 \sigma_P^2 \quad \text{ja} \quad \sigma_{A_o P_d} = \sigma_{A_o P_s} = \frac{1}{2} \sigma_A^2 = \frac{1}{2} h^2 \sigma_P^2.$$

Seega on kaaluparameetrid leitavad maatriksvõrdusest

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} \sigma_P^2 & \frac{1}{2} h^2 \sigma_P^2 & \frac{1}{2} h^2 \sigma_P^2 \\ \frac{1}{2} h^2 \sigma_P^2 & \sigma_P^2 & 0 \\ \frac{1}{2} h^2 \sigma_P^2 & 0 & \sigma_P^2 \end{pmatrix}^{-1} \begin{pmatrix} h^2 \sigma_P^2 \\ \frac{1}{2} h^2 \sigma_P^2 \\ \frac{1}{2} h^2 \sigma_P^2 \end{pmatrix}. \quad (5.25)$$

Näide 5.6. Uuritavaks tunnuseks on tallede 100-päeva kehamass päritavusega $h^2 = 0,3$ ja fenotüübidispersiooniga $\sigma_P^2 = 84,3$. Tall, kellele tahame hinnata aretusväärtust, kaalus 100-päevaselt 36 kg, tema isa 46 kg ja ema 34 kg. Populatsiooni keskmine tallede 100-päeva kehamass oli 30,3 kg.

Selektiooniindeksi (5.24) kordajate leidmiseks kirjutame välja maatriksvõrduse (5.25):

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 84,3 & 12,645 & 12,645 \\ 12,645 & 84,3 & 0 \\ 12,645 & 0 & 84,3 \end{pmatrix}^{-1} \begin{pmatrix} 25,29 \\ 12,645 \\ 12,645 \end{pmatrix} = \begin{pmatrix} 0,27 \\ 0,11 \\ 0,11 \end{pmatrix},$$

millest analüüsitava talle 100-päeva kehamassi aretusväärtuseks saame

$$\hat{A}_o = 0,27(36 - 30,3) + 0,11(46 - 30,3) + 0,11(34 - 30,3) = 3,65.$$

5.4 KASUMIINDEKSID

Reaalse aretuse eesmärk on loomapidamisest saadava majandusliku kasumi maksimaalne suurendamine. Selleks ei piisa aga üksnes fenotüübiväärtustel ja populatsiooni geneetilistel parameetritel baseeruvast valikust, lisaks on vaja arvestada ka iga selektsiooni all oleva tunnuse ühiku võrra parandamisega kaasneva majandusliku efektiga. Selliseid, iga üldises selektsiooniindeksis sisalduva tunnuse muutmise majanduslikule skaalale üleviimiseks kasutatavaid täiendavaid kaaluparameetreid nimetatakse **majanduslikeks kaaludeks**. Et kogu aretus on suunatud tulevikku, on ka majanduslike kaalude määramisel vajalik osata ette näha majanduslikku (ja ka poliitilist) situatsiooni tulevikus, samuti on hädavajalik omada võimalikult täpseid andmeid loomakasvatusega kaasnevate kulude ja tulude kohta.

Üldine valem **kasumiindeksi** (nimetatud ka kui **agregaatgenotüüp**, inglise keeles *aggregate genotype*) arvutamiseks on kujul

$$H = v_1A_1 + v_2A_2 + \dots + v_mA_m,$$

kus A_i tähistab tunnuse i aretusväärtust ja v_i selle tunnuse majanduslikku kaalu.

Näide 5.7. Olgu eesti holsteini tõugu lehma Lehvi piima-, rasva- ja valgu 1. laktatsiooni toodangu aretusväärtused vastavalt +600 kg, +10 kg ja +15 kg.

Vastavalt aretusprogrammile on eesmärgiks suurendada eelkõige piima valgu- ja rasvasisaldust, püüdes seejuures jätta kogutoodangu mahu vähemalt samale tasemele. Sellest lähtuvalt võetakse valgutoodangu majanduslikuks kaaluks +4, rasvatoodangu majanduslikuks kaaluks +1 ja piimatoodangu majanduslikuks kaaluks 0.

Lehvi kasumiindeksi väärtuseks (ehk suhteliseks aretusväärtuseks) saame:

$$I = 0 \times 600 + 1 \times 10 + 4 \times 15 = +70.$$

5.5 SELEKTSIOONIINDEKSITE TEISENDAMINE

Tavapäraselt avaldatakse indekseid väärtused mingi fikseeritud keskväärtuse suhtes, andes seejuures ette ka indeksi standardhälbe (keskmise erinevuse keskväärtusest) – saadud indekseid (aretusväärtuseid) nimetatakse **suhtelisteks (jõudluse, välimiku, viljakuse, ...) indeksiteks (aretusväärtusteks)**. Keskmise indeksi väärtus võetakse enamasti võrdseks 100-ga, standardhälve on aga erinevates riikides ja erinevate loomaliikide puhul erinev (näiteks eesti hobuste paremusjärjestusse panekul 20 ja JKK poolt hinnataval lüpsikarja suhtelise piimajõudluse aretusväärtusel 12). Etteantud keskmise ja standardhälbe suhteline indeks I_S (või suhteline aretusväärtus SAV) saadakse esialgselt indeksist I järgmise valemi abil:

$$I_S = \frac{I - \bar{I}}{\sigma_I} \times \sigma_{I_S} + \bar{I}_S,$$

kus \bar{I} ja \bar{I}_S tähistavad vastavalt esialgse ja uue indeksi keskväärtusi ning σ_I ja σ_{I_S} standardhälbeid³.

Näide 5.8. Jätkame eelmise punkti näidet ja oletame, et korrigeerimata indekseid keskväärtus on 0 ja standardhälve 50. Kehtestame uueks keskmiseks 100 ja standardhälbeks 12 punkti. Lehm Lehvi indeksi väärtus +70 teiseneb siis tulemuseks

$$I_S = \frac{I - \bar{I}}{\sigma_I} \times \sigma_{I_S} + \bar{I}_S = \frac{70 - 0}{50} \times 12 + 100 = 116,8 \text{ punkti.}$$

³ Kogu uuritava populatsiooni keskmise indeksi \bar{I} asemel leitakse suhteline indeks sageli mingi kindla loomade rühmi (näiteks hindamishetkest 10-12 aastat tagasi sündinud või kõigi mitteteadaolevate vanematega loomade) keskmise indeksi (aretusväärtuse) suhtes.

5.6 ÜLESANDED

1. Tunnuse päritavus on 0,3 ja korduvus 0,4. Kas selektsioon, mis baseerub antud tunnuse kahekordisel määramisel samal loomal, on rohkem või vähem täpne võrreldes seitsmel järglasel (poolõed) määratud väärtustel baseeruva selektsiooniga?
2. Lähtudes valemist (5.12) leia aretusväärtused tabelis 1 toodud eesti tõugu täkkudele nende järglaste hindamistulemuste alusel eraldi kõigile tunnustele.
 - Arvuta igale täkule kasumiindeksi väärtus võttes aluseks tabelis 2 toodud majanduslikud kaalud.
 - Teisenda kasumiindeksi väärtuseid nii, et nende keskmine oleks 100 ja standardhälve 20, ning järjestä täkkude saadud indeksiväärtuste alusel.
 - Kas oskate välja pakkuda sellise majanduslike kaalude kombinatsiooni, mille korral oma majandusgeneetiliselt potentsiaalilt esimene ja viimane täkk vahetaksid kohad?
 - Hinnake täkkudele aretusväärtused, kasutades kordaja 2 asemel valemiga (5.18) määratud kordajat. Lihtsuse mõttes võib kõigi hinnatud näitajate päritavuse võtta võrdseks 0,25-ga. Kas täkkude aretusväärtused ja paremusjärjestus muutuvad?

PS. Tulemused võite vormistada tabeli 3 kujul.

Tabel 1. Viie eesti tõugu täku järglaste hindamistulemused 2000. aastast

Hobuse nimi	Isa	Tõutüüp	Kehaehitus	Jalad	Samm	Traav	Vabahüpe
Vaida	Vigur 682 E	7	8	7	7	7	8
Vaara	Vigur 682 E	7	8	7	6	6	7
Villu	Vigur 682 E	7	7	6	8	6	6
Viker	Vigur 682 E	6	6	6	7	7	5
Teini	Tukker 703 E	8	9	7	8	7	8
Taara	Tukker 703 E	7	8	6	7	6	7
Tenor	Tiktor 697 E	6	6	7	6	7	6
Tommy	Tiktor 697 E	7	7	7	6	6	7
Relli	Rosett 600 E	7	7	7	7	7	9
Rollu	Rosett 600 E	6	7	8	7	6	6
Raul	Rommik 710 E	8	8	7	7	6	7
Rolf	Rommik 710 E	9	8	8	7	7	7
Rubik	Rommik 710 E	7	6	5	6	5	6

Tabel 2. Tunnuste majanduslikud kaalud

	Tõutüüp	Kehaehitus	Jalad	Samm	Traav	Vabahüpe
Majanduslik kaal	1	1	1,5	2	2	2,5
Majanduslik kaal (2)

Tabel 3. Tulemuste tabel

		Täkuld				Üld- keskmine	Standard- hälve
		Vigur	Tukker	Tiktor	Rosett		
Järglaste keskmine	Tõutüüp						
	Kehaehitus						
	Jalad						
	Samm						
	Traav						
	Vabahüpe						
Aretus- väärtus	Tõutüüp						
	Kehaehitus						
	Jalad						
	Samm						
	Traav						
	Vabahüpe						
Kasumiindeks							
Teisend. kasumiindeks						100	20
Jrk. nr.							
Kasumiindeks (2)							
Teisend. kasumiindeks (2)						100	20
Jrk. nr. (2)							
Päritavus $h^2 = 0,25$							
Järglaste arv n							
Aretus- väärtus [vale- mi (5.18) abil]	Tõutüüp						
	Kehaehitus						
	Jalad						
	Samm						
	Traav						
	Vabahüpe						
Kasumiindeks							
Teisend. kasumiindeks						100	20
Jrk. nr.							