II

POPULATSIOONIGENEETIKA FENOTÜÜPIDE TASEMEL

2.1 Polügeensed tunnused

Suurem osa morfomeetrilisi (arvulisi, mõõdetavaid) tunnuseid (keha mõõtmed, produktsiooninäitajad jms), mis iseloomustavad fenotüüpi, sõltuvad samaaegselt paljudest geenidest, st on polügeensed. Polügeensete tunnuste geneetika käsitleb tunnust määravate geenide summat ühtse tervikkomplektina (genotüübina), laskumata üksikute geenide tasemele.

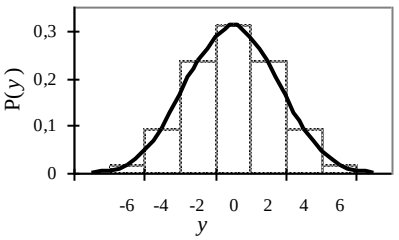
Polügeensete tunnuste väärtus on juhuslik, sõltudes ühelt poolt geenide konkreetsest komplektist genotüübis, nende omavahelisest paigutusest ja koosmõjust (dominantsi- ja epistaasinähtudest jmt), teisalt ka keskkonnatingimustest, milles organism on kasvanud. Sõltumine paljudest faktoritest on põhjuseks, miks polügeensed tunnused jaotuvad sageli vastavalt normaaljaotusele.

Näide. Oletame näiteks, et kehapikkus *y* (fenotüüp[1]) sõltub kolmest geenist (tegelikkuses on pikkust mõjutavate geenide arv kindlasti palju suurem), ja olgu esimeses, teises ja kolmandas lookuses dialleelsed süsteemid vastavalt alleelidega *A* ja *a*, *B* ja *b* ning *C* ja *c*, igas lookuses kumbki alleel võrdse tõenäosusega. Oletame, et väiketähega tähistatud alleeli esinemine vähendab indiviidi kehapikkust 1 ühiku võrra, suurtähega tähistatud alleelid aga suurendavad samapalju. Keskkonna mõju jätame esialgu vaatlusest välja. Siis saame Hardy-Weinbergi tasakaalu korral järgmise tabeli ja sellele vastava graafiku, kus *y* on indiviidi pikkus ja P(*y*) on pikkusega *y* indiviidide osakaal populatsioonis.

Tabel 2.1. Kolme dialleelse lookuse alusel moodustatavad genotüübid, neile vastavad fenotüübid ja osakaalud populatsioonis.

Genotüübid (*Aa* ja *aA* jne on kirjas ühe variandina)	*y*	P(*y*)
AABBCC	6	1/64
AABBCc, AABbCC, AaBBCC	4	6/64
AABBcc, AABbCc, AaBBCc, AAbbCC, AaBbCC, aaBBCC	2	15/64
AABbcc, AaBBcc, AAbbCc, AaBbCc, aaBBCc, AabbCC, aaBbCC	0	20/64
AAbbcc, AaBbcc, AabbCc, aaBBcc, aaBbCc, aabbCC	-2	15/64
Aabbcc, aaBbcc, aabbCc	-4	6/64
aabbcc	-6	1/64

Graafikult on näha, et pikkuse jaotus populatsioonis on lähedane normaaljaotusele. Siin avaldub tsentraalne piirteoreem, mille kohaselt binoomjaotus läheneb binoomi astme suurenemisel normaaljaotusele. Normaaljaotus on tüüpiline morfomeetrilistele tunnustele, mille väärtust kujundavad paljud vähesõltuvad või sõltumatud faktorid, millest ükski ei domineeri oluliselt teiste üle.



Joonis 2.1. Binoomjaotuse lähendamine normaaljaotusega.

---

[1] **Fenotüüp** (*phenotype*) – indiviidi (morfoloogiliste, füsioloogiliste, keemiliste, etoloogiliste, arenguliste) tunnuste (variantide ja avaldumistasemete) vaadeldav kogum; kitsamas mõistes üksiku uuritava tunnuse väärtus. **Genotüüp** (*genotype*) – 1) indiviidi (või raku) kogu geneetiline informatsioon, mis koostoimes keskkonnatingimustega määrab tema fenotüübi; 2) indiviidi (raku) geneetiliste lookuste alleelne koosseis.

2.2 Polügeense tunnuse väärtuste modelleerimine

2.2.1 Mudeli põhikuju

Polügeense tunnuse väärtuste modelleerimisel võib edukalt kasutada **üldisi lineaarseid mudeleid**. Nende mudelite kohaselt kujuneb tunnuse väärtus tunnuse teatava standardväärtuse (enamasti keskväärtuse) baasil, mida modifitseerivad (täpsustavad, muudavad) mitmesugused fikseeritud ja juhuslikud **faktorid**[2]. Matemaatiliselt esitatakse lineaarne mudel diskreetsete[3] faktorite väärtustele (**tasemetele** e **nivoodele**) vastavate arvväärtuste (**mõjude** e **efektide**) ning sobivate kordajatega pidevate argumenttunnuste väärtuste lineaarse funktsioonina.

Lihtsaim populatsioonigeneetikas kasutatav mudel esitab indiviidi fenotüübiväärtuse *y* populatsiooni keskmise fenotüübiväärtuse μ, genotüübi mõju *G* ja keskkonnamõjude *E* summana:

$$y = \mu + G + E\,. \tag{2.1}$$

Teoreetilistes uuringutes võib keskkonnamõjud jätta sageli arvesse võtmata, mistõttu punktis 2.1 toodud näite korral sobiksid fenotüübiväärtuse modelleerimiseks teiste hulgas järgmised lineaarsed mudelid:

$$y_{ijklmn} = \mu + G1_i + G2_j + G3_k + G4_l + G5_m + G6_n\,, \tag{2.2}$$

$$y_{ijk} = \mu + L1_i + L2_j + L3_k\,, \tag{2.3}$$

$$y_i = \mu + G_i\,. \tag{2.4}$$

Siin *y* on tunnuse väärtus, μ on tunnuse „standardväärtus" ja ülejäänud liikmed vastavad geneetilistele faktoritele. Indeksid näitavad geneetiliste faktorite väärtusi (e nivoosid, tasemeid) antud indiviidi korral. Mudelis (2.2) on polügeenset tunnust *y* mõjutavateks geneetilisteks faktoriteks 6 geeni, mudelis (2.3) 3 lookust ja mudelis (2.4) üks genotüüp. Igal geenil on 2 võimalikku väärtust (näiteks esimese geeni väärtusteks mudelis (2.2) on *A* ja *a*, teise geeni väärtusteks samuti *A* ja *a*, kolmanda geeni väärtused on *B* ja *b* jne). Esimese lookuse väärtused mudelis (2.3) on *AA*, *Aa* ja *aa* (eeldusel, et juhte *Aa* ja *aA* ei eristata). Genotüübi väärtused (kui vastava faktori tasemed) mudelis (2.4) on toodud tabelis 2.1 (27 erinevat väärtust). Iga faktori igale väärtusele vastab mudelis arvuline term. Näiteks mudelis (2.2) on esimese geeni nivoodele vastavateks efektideks (mõjudeks) $G1_1 = 1$ ja $G1_2 = -1$. Edaspidi nimetame neid väärtusi **geeniväärtusteks**. Mudelis (2.3) on esimese lookuse nivoodele vastavateks efektideks $L1_1 = 2$, $L1_2 = 0$ ja $L1_3 = -2$.

Indeksitega saab iga tunnuse väärtuse (ja indiviidi, kellel see väärtus on mõõdetud) siduda vastavate faktorite väärtustega (antud juhul indiviidi geneetilise konstitutsiooniga). Näiteks mudelis (2.3) tähistab y_{231} tunnuse väärtust indiviidil, kelle genotüüp on *AabbCC* ja seega $y_{231} = \mu + L1_2 + L2_3 + L3_1 = 0 + 0 + (-2) + 2 = 0$.

Üldiste lineaarsete mudelite teoorias kasutatakse üksikasjaliste mudelispetsiifiliste esituste asemel märksa universaalsemat ja lühemat maatrikskuju

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}\,. \tag{2.5}$$

Selles esituses on **y** vaatluste vektor, mille moodustavad (tavaliselt indeksite järgi leksikograafiliselt järjestatud) tunnuse *y* väärtused geneetiliste faktorite kõikvõimalike erinevate kombinatsioonide juures, näiteks mudeli (2.2) korral oleks vektoril **y** 64 komponenti, vastavalt indeksitele (111111), (111112), …, (222222); **β** on kõigi mudelis sisalduvate efektide vektor e nn parameetermaatriks, mille

---

[2] **Fikseeritud faktoril** on: a) vähe nivoosid, b) iga nivoo pakub iseseisvat huvi ja on valitud mittejuhuslikult, c) andmetes on või saavad põhimõtteliselt olla esindatud kõik nivood. Seetõttu käsitletakse fikseeritud faktorite tasemetele vastavaid efekte kui konstante.
Juhuslikul faktoril on a) potentsiaalselt väga palju (lõpmatu hulk) nivoosid, b) andmetes on neist esindatud juhuslik valim, c) huvi pakub kõigi (ka andmetes esindamata) nivoode keskmine mõju e see, kui suur osa uuritava tunnuse koguvarieeruvusest on kirjeldatud antud faktori poolt. Juhuslike faktorite tasemetele vastavaid efekte käsitletakse kui mingi teoreetilise jaotusega juhuslike suuruste realiseerunud väärtusi, kus selle teoreetilise jaotuse näol mõistetakse üldjuhul normaaljaotust.

[3] **Diskreetne argumenttunnus** e **faktortunnus** ei pruugi olla arvuline, omab lõpliku hulga väärtusi ja kirjeldab funktsioontunnust oma tasemetele vastavate arvuliste mõjude kaudu.
Pidev argumenttunnus on arvuliste väärtustega ja kirjeldab funktsioontunnuse väärtusi läbi mingi funktsiooni oma väärtustest.
Üksnes pidevate ja omavahel sõltumatute argumenttunnustega mudel on **regressioonanalüüsi mudel**, üksnes diskreetsete ja omavahel sõltumatute argumenttunnustega mudel on **dispersioonanalüüsi mudel**.

moodustavad lineaarse mudeli kõikvõimalikud liikmed (parameetrid) alates standardväärtusest μ (näiteks mudeli (2.2) korral $\boldsymbol{\beta}^{\mathrm{T}} = (\mu\ G1_1\ G1_2\ G2_1 \cdots G3_2)$); **X** on 0-dest ja 1-dest koosnev (diskreetsete faktorite puhul) **plaani**- e **disainimaatriks** (ka mudelimaatriks), mille ridade arv võrdub vaatluste arvuga ja veergude arv efektide arvuga ning mis seostab iga vaatluse just temale vastavate faktorite tasemetega. Näiteks mudel (2.3) esitub järgmise maatriksvõrdusena (eeldame, et igale genotüübile vastab vaid üks vaatlus ning konkreetse lookuse osas heterosügootsed genotüübid on fenotüübilt identsed ja esindatud samuti vaid ühe vaatlusega):

$$\underset{\mathbf{y}_{27\times1}}{\begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ \vdots \\ y_{333} \end{pmatrix}} = \underset{\mathbf{X}_{27\times10}}{\begin{pmatrix} 1&1&0&0&1&0&0&1&0&0 \\ 1&1&0&0&1&0&0&0&1&0 \\ 1&1&0&0&1&0&0&0&0&1 \\ 1&1&0&0&0&1&0&1&0&0 \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ 1&0&0&1&0&0&1&0&0&1 \end{pmatrix}} \underset{\boldsymbol{\beta}_{10\times1}}{\begin{pmatrix} \mu \\ L1_1 \\ L1_2 \\ L1_3 \\ \vdots \\ L3_3 \end{pmatrix}} = \begin{pmatrix} \mu + L1_1 + L2_1 + L3_1 \\ \mu + L1_1 + L2_1 + L3_2 \\ \mu + L1_1 + L2_1 + L3_3 \\ \mu + L1_1 + L2_2 + L3_1 \\ \vdots \\ \mu + L1_3 + L2_3 + L3_3 \end{pmatrix}.$$

Juhul, kui lineaarseid mudeleid kasutatakse reaalsete andmete analüüsil, püüdes kirjeldada uuritava tunnuse väärtusi mingite faktorite mõju abil (regressioonanalüüs, dispersioonanalüüs), on loomulik, et mudel ei ole 100% täpne ja alati jääb mingi osa uuritava tunnuse varieeruvusest konstrueeritud mudeli abil kirjeldamata. Taolist mudeli poolt kirjeldamata jäänud osa uuritava tunnuse väärtustest nimetatakse **mudeli veaks**, tähistatakse seda enamasti tähega e (või ε) ning esitatakse mudelis täiendava liikmena:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{2.6}$$

kus vektor **e** sisaldab igale vaatlusele vastavat mudeli poolt kirjeldamata jäänud osa ($\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$). Geneetikas loetakse mudeli viga sageli kuuluvaks keskkonnamõjude hulka ja eraldi liidetavana mudelis ära ei tooda.

2.2.2 Geneetilised interaktsioonid

Mudelid (2.2)-(2.4) on samaväärsed vaid siis, kui interaktsioonid geenide vahel puuduvad. Interaktsioon sama lookuse geenide (alleelide) vahel on **dominants**, interaktsioon eri lookuste geenide vahel on **epistaas** (joonis 2.2).

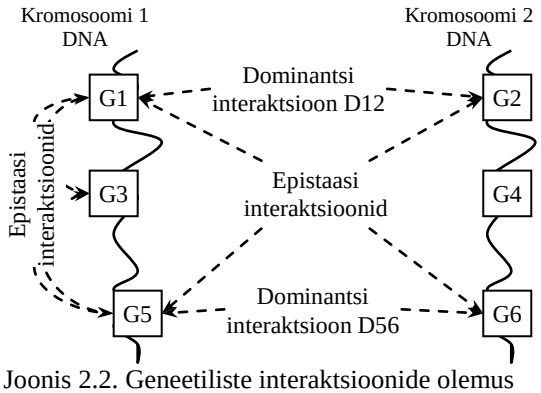
Vaatame näitena mudelit (2.2), mida on täiendatud dominantsi kirjeldavate interaktsiooniliikmetega $D12$ (kirjeldab esimese lookuse alleelide omavahelist dominatsisuhet), $D34$ ja $D56$ (vastavalt teise ja kolmanda lookuse alleelide dominants):

$$y_{ijklm} = \mu + G1_i + G2_j + G3_k + G4_l + G5_m + G6_n + D12_{ij} + D34_{kl} + D56_{mn}. \tag{2.7}$$

Osutub, et mudel (2.3) on saadav mudelist (2.7), kui defineerida

$$L1_{ij} = G1_i + G2_j + D12_{ij},\ L2_{kl} = G3_k + G4_l + D34_{kl} \text{ ja } L3_{mn} = G5_m + G6_n + D56_{mn}.$$

Täiendades mudelit (2.7) interaktsioonidega eri lookuste geenide vahel (epistaasid), saaksime mudeli, mis omadustelt läheneb mudelile (2.4), kuid jääb ikkagi „jämedamaks" selles mõttes, et genotüübi kogumõju kirjeldamiseks ei piisa dominantsi- ja epistaasiinterakstioonidest.



Joonis 2.2. Geneetiliste interaktsioonide olemus

2.2.3 Mudeli reparametriseerimine

Geneetilise mudeli parameetrid on määratud võrrandist (2.5) tuleneva seosega[4]

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} . \tag{2.8}$$

Võrrandi (2.8) alusel on mudel üheselt mõtestatav (ühetähenduslik), kui leidub pöördmaatriks $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$, milleks peab plaanimaatriks **X** olema täisveeruastakuga. Viimane tingimus tähendab plaanimaatriksi **X** kõigi veergude lineaarset sõltumatust, mis aga ei kehti ühegi vaadeldud geneetilise mudeli korral (ühe ja sama geeni, lookuse või genotüübi erinevatele tasemetele vastavad plaanimaatriksi veerud annavad kokku liites „standardväärtusele" vastava ühtede veeru).

Garanteerimaks mudeli ja sellega kirjeldatud fenotüübi ja genotüübi vahelise seose ühetähenduslikkust, peab tegema teatavad lisakitsendused mudeli parameetrite kohta (e mudel peab olema sobivalt parametriseeritud). Traditsiooniliseimad nn **reparametriseerimistingimused** nõuavad, et fikseeritud faktori nivoodele vastavate efektide summa peab olema null, juhusliku faktori korral peab efekti (juhusliku suuruse) keskväärtus olema null. Reparametriseerimistingimused on alati rahuldatavad mudeli üldisust kitsendamata. Näiteks mudeli (2.3) korral on $L1_1 + L1_2 + L1_3 = 0$, $L2_1 + L2_2 + L2_3 = 0$ ja $L3_1 + L3_2 + L3_3 = 0$, mudeli (2.4) korral $G_1 + ... + G_{27} = 6 + 4 + 4 + 4 + 2 + ... - 6 = 0$.

Reparametriseerimine geneetilisi interaktsioone modelleerivates mudelites toimub analoogsete tingimustega. Näiteks mudeli (2.7) puhul on loomulik nõuda, et kehtiks võrdused
$\sum_{i=1}^{2} D12_{ij} = 0, (\forall j)$, $\sum_{j=1}^{2} D12_{ij} = 0, (\forall i)$, $\sum_{k=1}^{2} D34_{kl} = 0, (\forall l)$, $\sum_{l=1}^{2} D34_{kl} = 0, (\forall k)$, $\sum_{m=1}^{2} D56_{mn} = 0, (\forall n)$
ja $\sum_{n=1}^{2} D56_{mn} = 0, (\forall m)$.

Üldiste lineaarsete mudelite maatriksesituses määratakse reparametriseerimistingimused 0-dest ja 1-dest koosneva maatriksi **H** abil kujul

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{0} . \tag{2.9}$$

Kusjuures maatriks **H** konstrueeritakse nii, et

$$\operatorname{rank}\begin{pmatrix}\mathbf{X}\\ \mathbf{H}\end{pmatrix} = \operatorname{rank}(\mathbf{X}) + \operatorname{rank}(\mathbf{H}) = \mathbf{X}\text{-i veergude arv} . \tag{2.10}$$

Geneetilise mudeli parameetreid määrav võrrand (2.8) teiseneb siis kujule

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} . \tag{2.11}$$

Näide. Oletame, et fenotüübiline tunnus *y* sõltub ühest dialleelsest lookusest (kahest geenist, millel mõlemal on kaks esinemisvarianti, näiteks *A* ja *a*) vastavalt mudelile

$$y_{ij} = \mu + G1_i + G2_j + D12_{ij}$$

ning olgu tunnuse väärtused erinevate genotüüpide korral järgmised: $y_{11} = 28$ (genotüüp *AA*), $y_{12} = 15$ (*Aa*), $y_{21} = 15$ (*aA*), $y_{22} = 12$ (*aa*), seega vaatluste vektor $\mathbf{y} = (28\ 15\ 15\ 12)^{\mathrm{T}}$. Plaanimaatriks **X**, mis vastavalt mudelile (2.5) seob fenotüübiväärtused geeniväärtuste ja dominantsiefektidega, on kujul

$$\mathbf{X} = \begin{array}{c} \\ AA \\ Aa \\ aA \\ aa \end{array}\!\!\begin{array}{c} \begin{array}{cccccccccc} \mu & G1_1 & G1_2 & G2_1 & G2_2 & D12_{11} & D12_{12} & D12_{21} & D12_{22} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} .$$

Maatriksi **X** astak on 4 (kontrolli!). Mudeli parameetrite üheseks määramiseks vajalikud reparametriseerimistingimused ja nende maatriksvõrdusesse (2.11) kaasamiseks tarvilik maatriks **H** on kujul

---

[4] Mudelist (2.6) hinnatakse parameetritevektor $\boldsymbol{\beta}$ vähimruutude meetodil, st et parameetervektori $\boldsymbol{\beta}$ hinnang $\hat{\boldsymbol{\beta}}$ valitakse selliselt, et mudeli vigade ruudud oleks minimaalsed. Maatrikskujul on vähimruutude tingimus mudeli (2.6) tarvis väljendatav seosena $\min_{\boldsymbol{\beta}}(\mathbf{e}^{\mathrm{T}}\mathbf{e}) = \min_{\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.
Hinnangu $\hat{\boldsymbol{\beta}}$ avaldamiseks tuleb mudeli (2.6) vigade ruutude summast

$$\mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$$

võtta tuletis $\boldsymbol{\beta}$ järgi ja võrdsustada tulemus nulliga. Diferentseerimise tagajärjel saame avaldise $\partial \mathbf{e}^{\mathrm{T}}\mathbf{e} / \partial \boldsymbol{\beta} = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$, mille nulliga võrdsustamisest järeldub, et $\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$, millest omakorda $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$.

$$\mathbf{H} = \begin{array}{c} \\ G1_1 + G1_2 = 0 \\ G2_1 + G2_2 = 0 \\ D12_{11} + D12_{12} = 0 \\ D12_{11} + D12_{21} = 0 \\ D12_{12} + D12_{22} = 0 \\ D12_{21} + D12_{22} = 0 \end{array} \begin{array}{c} \begin{array}{cccccccccc} \mu & G1_1 & G1_2 & G2_1 & G2_2 & D12_{11} & D12_{12} & D12_{21} & D12_{22} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}.$$

Kokkuvõttes saame võrrandist (2.11) mudeli parameetrite väärtusteks

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ G1_1 \\ G1_2 \\ G2_1 \\ G2_2 \\ D12_{11} \\ D12_{12} \\ D12_{21} \\ D12_{22} \end{pmatrix} = \begin{pmatrix} 17{,}5 \\ 4 \\ -4 \\ 4 \\ -4 \\ 2{,}5 \\ -2{,}5 \\ -2{,}5 \\ 2{,}5 \end{pmatrix}.$$

Ülesanne 12.
a) Veenduda, et maatriks **H** rahuldab reparametriseerimistingimusi (2.9) ja (2.10) ning kehtib võrdus (2.5).
b) Uurida, kuidas muutuvad tulemused ja mudel, kui jätta ära parameeter $D12_{21}$, mis on võrdne parameetriga $D12_{12}$.
c) Leida parameetrite väärtused valemi (2.8) alusel, kasutades mudeli alternatiivset ja näiteks statistikapaketis SAS kasutatavat reparametriseerimist – viimase kohaselt kitsendatakse plaanimaatriksit, jättes sealt välja iga faktori viimasele tasemele vastavad veerud ja ka kõik ülejäänud lineaarselt sõltuvad veerud (misläbi ka vastavad faktorite tasemete väärtused võrdsustatakse 0-ga) ning lahendatakse võrrand kitsendatud plaanimaatriksi abil vaid mittenulliliste parameetrite suhtes – ning veenduda, et ka sel juhul kehtib võrdus (2.5).

2.3 Geneetilised parameetrid ja nende hindamine

2.3.1 Aretusväärtus

Indiviidi **aretusväärtus** (*breeding value*) ehk aditiivne geneetiline väärtus on kõigi indiviidi genotüübis olevate alleelide summaarne efekt (geeniväärtuste summa). Nimetus tuleneb sellest, et põllumajanduslikus aretustöös (mille tarvis populatsioonigeneetikas kasutatavad lineaarsed mudelid esmalt välja töötati) on tähtsus vaid tunnuse jaoks „soodsate" alleelide summaarsel mõjul. Milline on konkreetne genotüüp (alleelide paar) mingis lookuses või üleüldse, missuguses lookuses need geenid on, ei ole oluline, sest esiteks pärandub järglasele igast lookusest vaid üks alleel ja teiseks toimub ristumiste käigus geenide ümberpaiknemine, mistõttu interaktsioonide efektid ei ole püsivad ja aretuse seisukohast tähtsust ei oma.

Teoreetilistes uuringutes, kus on täpselt teada (ära fikseeritud) nii tunnust määravate geenide arv, geeni- ja genotüübiväärtused kui ka indiviidide genotüübid, on tegu nn **teoreetiliste aretusväärtustega**. Näiteks mudelist (2.7) avalduvad teoreetilised aretusväärtused summana $G1_i + G2_j + G3_k + G4_l + G5_m + G6_n$.

Praktilistes populatsioonipõhistes uuringutes ei ole tunnust määravate geenide arv, geeni- ja genotüübiväärtused ja uuritavate indiviidide genotüübid teada ning lisaks ei pruugi reaalses populatsioonis kõiki teoreetiliselt eksisteerida võivaid alleele üleüldse leidudagi. Seetõttu kasutatakse tegelike andmetega töötamisel **hinnatud aretusväärtust**, mille leidmisel võetakse aluseks lineaarne mudel kujul

$$y_i = \mu + A_i + I_i + E_i \,, \tag{2.12}$$

kus y_i on *i*-nda indiviidi fenotüübiväärtus, μ on ligikaudselt väljendudes vaatlusaluse populatsiooni keskmine fenotüübiväärtus (mitte aga kõigi teoreetiliselt võimalike genotüüpide keskmine nagu teoreetilise aretusväärtuse arvutamisel aluseks olevais valemeis), A_i on *i*-nda indiviidi aretusväärtus, I_i on interaktsiooniefekt (dominantsi- ja epistaasiefektid kokku) ja E_i kirjeldab keskkonnatingimuste mõju (mille sisse kuuluvaks loetakse kogu mudelis sisaldunud geneetiliste efektide poolt kirjeldamata jäänud osa indiviidi *i* fenotüübiväärtusest). Efektid A_i, I_i ja E_i on juhuslikud suurused dispersioonidega vastavalt σ_A^2, σ_I^2 ja σ_E^2, kusjuures $A_i \perp I_i$, $A_i \perp E_i$, $I_i \perp E_i$, $A_i \perp A_{i'}$, $I_i \perp I_{i'}$ ja $E_i \perp E_{i'}$ ($i \neq i'$). Märgime, et nendest eeldustest kõige küsitavam on indiviidi aretusväärtuse ja talle mõjuvate keskkonnamõjude sõltumatus ($A_i \perp E_i$), sest paremad (suurema aretusväärtusega A_i) indiviidid satuvad sageli paremasse keskkonda (neid hooldatakse paremini nende paremate omaduste tõttu).

Mudel (2.12) ei ole kahjuks mõistlikult reparametriseeritav (reparametriseerimistingimused jätavad alles sisuliselt vaid ühe efektidest A_i, I_i või E_i, mistõttu neid efekte ei saa otseselt hinnata). Efekti A_i saab aga hinnata siis, kui on teada tunnuse *y* väärtused indiviidi *i* järglastel. Võtmeks on asjaolu, et indiviid annab järglasele pooled oma geenidest, seega keskmiselt poole oma geeniväärtuste summast ehk aretusväärtusest A_i. St, et indiviidi *i* järglase *j* fenotüübiväärtus y_{ij} on esitatav mudeliga

$$y_{ij} = \mu + \tfrac{1}{2} A_i + U_{ij} + V_{ij} + E_{ij} \,, \tag{2.13}$$

kus juhuslik suurus U_{ij} on indiviidi *i* järglase *j* teise vanema mõju, V_{ij} kirjeldab juhuslikkust (nn Mendeli valiku[5] mõju), mida tingib vanemate geenide juhuslik valik järglase *j* genotüüpi ja E_{ij} on kõigi ülejäänud faktorite juhuslik mõju järglase *j* tunnusele *y*. Kui eeldada, et indiviidi *i* järglased kasvavad juhuslikult valitud keskkondades ja nende teine vanem on samuti valitud juhuslikult, on juhuslike suuruste *U*, *V* ja *E* keskväärtused üle *j* nullid. Tähistades indiviidi *i* järglaste fenotüübiväärtuste keskväärtuse $\mathrm{E}_j(y_{ij}) = \bar{y}_{i\cdot}$, saame

$$\bar{y}_{i\cdot} = \bar{y}_{\cdot\cdot} + \tfrac{1}{2} A_i \,,$$

kus $\bar{y}_{\cdot\cdot} = \mu$ on tunnuse *y* keskmine üle lõpmatu populatsiooni. Siit tuleneb valem

$$A_i = 2(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \,, \tag{2.14}$$

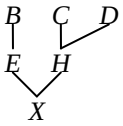
mis defineerib indiviidi aretusväärtuse kui kahekordse oodatava erinevuse tema järglaste ja populatsiooni keskmise vahel, eeldusel, et teine vanem on valitud populatsioonist täiesti juhuslikult ning geenid ei ole aheldunud.

Aretusväärtuse alternatiivne definitsioon esitataksegi sageli just viimase valemi sõnastusena, kusjuures võtmesõnadeks on siin

- lõpmatu populatsioon ja lõpmatu hulk järglasi ning
- teise vanema juhuslik valik ja seeläbi populatsiooni keskmisele vastavus.

[5] Mendeli valikuks nimetatakse meioosi protsessis tehtavat lookusesisest alleelide vahelist valikut – valitakse, kumb alleelidest järglasele pärandub.

Näide. Olgu indiviidid seotud juuresoleval joonisel kujutatud sugulussidemetega. Olgu teadaolevad aretusväärtused järgmised: $A_B = 0{,}5$, $A_C = -0{,}5$ ja $A_D = 1$ ning populatsiooni keskmise loeme võrdseks nulliga. Skeemil näitamata indiviidi *E* teine vanem on valitud populatsioonist huupi. Prognoosime A_X.



Kõigepealt leiame $A_E = ½(A_B + 0) = 0{,}25$, võttes indiviidi *E* tundmatu vanema aretusväärtuseks 0, mis on populatsiooni keskmine. Teiseks leiame $A_H = ½(A_C + A_D) = 0{,}25$. Seejärel prognoosime $A_X = ½(A_E + A_H) = ½(0{,}25 + 0{,}25) = 0{,}25$. Seega on tunnuse oodatav väärtus indiviidil *X* umbes 0,25 ühiku võrra suurem populatsiooni keskmisest.

Valemi (2.14) puhul välistatud geenide aheldumine tähendab seda, et geenide edasikandumine järglastele ei ole sõltumatu. Sõltuvuse tavaliseks põhjuseks on geenide lähestikune paiknemine samas kromosoomis. Sel juhul on tõenäosus, et geenid satuvad ristsiirdes eri kromosoomidesse, väike ja järglased pärivad suure tõenäosusega ühe kahest geenikomplektist (haplotüübist), mis epistaasi tõttu võivad omada positiivset või negatiivset mõju järglastel mõõdetava tunnuse väärtustele. Valem (2.14) kirjutab aga erinevuse järglaste keskmise ja populatsiooni keskmise vahel aretusväärtuse arvele ka siis, kui selle erinevuse tingib epistaasiinteraktsioon.

Näide. Sõltugu tunnus kahest tugevasti aheldunud dialleelsest lookusest, kusjuures kõigi nelja geeni väärtused olgu $G1_1 = G1_2 = G2_1 = G2_2 = 0$ (siin on lühiduse mõttes esimeses lookuses paiknevate geenide võimalikud väärtused tähistatud *G*1 ja teises *G*2). Viimane tingimus tähendab kokkuvõttes, indiviidi *i* geeniväärtuste summa, so aretusväärtus, $A_i = 0$. Oletame edasi, et alleelid $G1_1$ ja $G2_1$ nii nagu ka alleelid $G1_2$ ja $G2_2$ on tugevasti aheldunud, kusjuures epistaasiinteraktsiooni efekt nii $G1_1$ ja $G2_1$ kui ka $G1_2$ ja $G2_2$ vahel on -1. Siis sõltumata sellest, kumma kromosoomi geenibloki (haplotüübi) järglane indiviidilt *i* pärib, on tunnuse väärtus temal ligikaudu 1 ühiku võrra väiksem populatsiooni keskmisest ja indiviidi *i* aretusväärtuseks saame valemi (2.14) alusel väära tulemuse $A_i = -2$.

Reaalseis polügeensete tunnuste populatsioonigeneetilistes analüüsides ei ole enamasti tarvidust muretseda võimaliku geenide ahelduse pärast, sest tuhandete geenide interaktsioonid tasakaalustavad üksteist.

Oluline on nentida, et erinevalt teoreetilisest aretusväärtusest sõltub reaalsete andmete alusel leitud aretusväärtus populatsiooni koostisest. Näiteks selektsiooni korral teoreetiline aretusväärtus ei muutu, reaalse populatsiooni põhine aretusväärtus muutub aga vastavalt populatsiooni struktuuri muutumisele aretustöös: mida homogeensem populatsioon, seda väiksem aretusväärtus. Seeläbi on oluline kasutada aretusväärtuse hindamisel populatsiooni esindavat valimit.

Et indiviidi aretusväärtuse näol on tegu juhusliku suurusega, mitte konstandiga, kerkib reaalsete populatsioonide uurimisel täiendav probleem aretusväärtuse dispersiooni (e **aditiivgeneetilise dispersiooni**) σ_A^2 hindamise näol.

- Aretusväärtuse hindamise ülesanne kerkib siis, kui otsitakse heade geneetiliste omadustega indiviidi, näiteks tõulooma,
- aditiivgeneetilise varieeruvuse hindamise probleem aga näiteks siis, kui soovitakse hinnata selektsiooni efektiivsust – kui σ_A^2 on väike, on geenide aditiivsest toimest tingitud erinevus populatsiooni keskmisest (aretusväärtus) kõigil indiviididel praktiliselt ühesugune ja selektsioon ei anna tulemusi.

Aretusväärtuse hindamisel annab lihtsamal juhul vastuse valem (2.14), kuid kui uuritava indiviidi järglaste arv on väike, on valemiga (2.14) saadud aretusväärtuse hinnang ebatäpne. Kui vaatlusalune populatsioon paikneb sarnastes keskkonnatingimustes, piisab täpsema hinnangu saamiseks indiviidi (ja/või tema lõpliku arvu sugulaste) fenotüübiväärtuste võrdlemisest populatsiooni keskmisega – taolist, kõigi nende fenotüübil mõõdetud erinevuste sobivalt valitud kordajatega kaalutuna ühte võrrandisse koondamist ja seeläbi aretusväärtuse hinnangu leidmist on käsitletud peatükis 2.3.4.

Juhul, kui uuritava populatsiooni geneetiline struktuur on liiga keeruline indiviididevaheliste sugulussidemete lihtsate võrranditega esitamiseks ja/või paiknevad indiviidid erinevates keskkonnatingimus-

tes, tuleb kasutusele võtta keerulisi kovariatsioonistruktuure modelleerida võimaldav **üldiste lineaarsete segamudelite teooria** (nimetatud ka kui **dispersioonanalüüsi segamudelite teooria**).

Aretusväärtuse dispersiooni arvutamiseks tuleb üldjuhul samuti kasutada üldisi lineaarseid segamudeleid. Lihtsamatel juhtudel piisab siiski ka tavalisest dispersioonanalüüsist[6].

2.3.2 Aretusväärtuse dispersiooni hindamine dispersioonanalüüsi abil

Vaatame esmalt lihtsuse mõttes eelmises punktis valemiga (2.13) modelleeritud situatsiooni, kus huvi pakub vaid ühelt vanemalt järglasele edasi kandunud geneetilise materjali mõju järglase fenotüübiväärtusele. Taolist mudelit on aastakümneid laialdaselt rakendatud põllumajandusloomade aretuses, kus huvipakkuvaks vanemaks on eelkõige isa – põhjuseks see, et ühel isal võib tänu kunstlikule seemendusele olla sadu ja isegi tuhandeid järglasi, kes elavad erinevais keskkonnatingimustes (mistõttu ühe isa järglaste keskmise keskkonnamõju võib lugeda võrdseks nulliga), ühe ema järglaste arv on reeglina tunduvalt väiksem ja sageli elavad järglased samades tingimustes (koos emaga), mistõttu ei osutu võimalikuks üheselt eristada emalt päritud geenide mõju ja keskkonnaefekti. Mudeli (2.13) esitus nn **isa mudelina** (*sire model*) on

$$y_{ij} = \mu + S_i + E_{ij}\,, \tag{2.15}$$

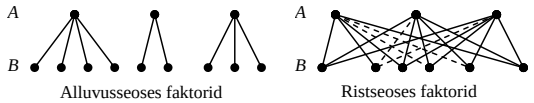
kus $S_i = \frac{1}{2} A_i$ (isa i mõju oma järglastele võrdub poolega tema aretusväärtusest) ja E_{ij} sisaldab kõiki ülejäänud (eeldatavalt isa mõjust sõltumatuid) geneetilisi ja mittegeneetilisi mõjusid. Eeldades, et genotüüp ja keskkond on sõltumatud ning vaatlusalune isa ega ükski tema järglaste emadest pole

---

[6] **Dispersioonanalüüs** (DA = ANOVA, *ANalysis Of VAriance*) on analüüsimeetod prognoosimaks ja kontrollimaks argumenttunnus(t)e mõju funktsioontunnusele. DA mudel on üldise lineaarse mudeli erijuht, kus eeldatakse, et 1) argumenttunnused on diskreetsed (nn faktorid e faktortunnused), mille tasemetele vastavate mõjude summana on avaldatav mistahes faktorite tasemete kombinatsioonile vastav uuritava tunnuse keskväärtus, 2) faktorite tasemete mõjud on sõltumatud. Viimasest eeldusest tuleneb, et uuritava tunnuse dispersioon on avaldatav faktorite mõjude kombinatsioonidele vastavate nn **dispersioonikomponentide** summana (siit ka nimetus dispersioonanalüüs).

Fikseeritud faktorite (vt allmärkus mõni leht tagasi) korral pakub huvi nende tasemetele vastavate mõjude hindamine ja ei oma mõtet dispersioonikomponendi mõiste (faktori mõju on täielikult määratud tema tasemetele vastavate arvväärtuste kaudu); juhuslike faktorite (kui juhuslike suuruste) mõju kirjeldab aga eelkõige nende panus uuritava tunnuse varieeruvusse läbi dispersioonikomponentide – mida suurem on mingi juhusliku faktori nivoodele vastavate mõjude varieeruvus (faktorile vastav dispersioonikomponent), seda enam on uuritava tunnuse varieeruvus määratud läbi selle faktori väärtuste, juhuslike faktorite üksiknivoode (kui juhusliku suuruse realiseerunud väärtuste) mõjude endi hindamine üldjuhul huvi ei paku (erandiks on siin eelkõige populatsioonigeneetika ja aretus, kus soovitakse lisaks dispersioonikomponentidele leida ka juhuslikena käsitletavate geneetiliste faktorite andmetes realiseerunud tasemete mõjude hinnanguid).

Vastavalt faktorite vahekorrale mudelis eristatakse **ristseoses** ja **alluvusseoses** olevaid faktoreid – esimesel juhul kombineerub (saab põhimõtteliselt kombineeruda) ühe faktori iga nivoo teise faktori kõigi nivoodega, teisel juhul aga esineb ühe faktori iga nivoo vaid koos konkreetse teise faktori nivooga. Seda, et faktor *B* allub faktorile *A*, tähistatakse enamasti kujul *B*(*A*). Lisaks võivad mudelis esineda faktorite interaktsioonid, näiteks kujul *AB*.



Dispersioonanalüüsi läbiviimise võib jagada järgmisteks sammudeks:

1) leitakse iga faktori mõjule vastav uuritava tunnuse varieeruvust kirjeldav suurus, nn ruutude summa (ruutvorm, *Sum of Squares*, *SS*),

2) leitakse faktorite tasemete arvust sõltuvad vabadusastmete (*degrees of freedom*, *df*) arvud,

3) arvutatakse faktorite mõjudele vastavad dispersiooni hinnangud (nn keskruudud, *Mean Squares*, *MS*) kujul $MS = SS/df$,

4) konstrueeritakse dispersiooni hinnangute keskväärtused E(*MS*) , mis on tundmatute dispersioonikomponentide funktsioonid,

5) lahendatakse saadud võrrand(isüsteem) dispersioonikomponentide suhtes,

6) konstrueeritakse sobivalt valitud (nullhüpoteesi alusel) keskruutude suhetena *F*-statistikud kontrollimaks hüpoteese faktorite mõjude olulisuse kohta (fikseeritud faktori korral väidab nullhüpotees faktori kõigi tasemete mõjude võrdumist nulliga, juhusliku faktori puhul aga vastava dispersioonikomponendi võrdumist nulliga),

7) arvutatakse soovi korral faktorite tasemete mõjud.

DA arvutused koondatakse DA tabelisse, kus igale faktorile vastab üks rida (ning lisaks on rida nii mudeli jäägi kui ka uuritava tunnuse koguvarieeruvuse tarvis) ja veergudes on arvutatud suurused *SS*, *df*, *MS*, E(*MS*), *F*, *p* (sõltuvalt analüüsi eesmärgist ja mudeli/faktorite olemusest ei pruugi olla leitud kõiki kolme viimasena nimetatut).

sugulased, on efektid S_i ja E_{ij} mudelis (2.15) sõltumatud ja uuritava tunnuse dispersioon (nn **fenotüübidispersioon**) avaldub summana

$$\sigma_y^2 = \sigma_S^2 + \sigma_E^2, \tag{2.16}$$

kus σ_S^2 väljendab isalt järglasele pärandunud geenide summaarse mõju dispersiooni ja σ_E^2 on isa mõjust kirjeldamata jäänud osa fenotüübilisest varieeruvusest (jääkvarieeruvus). Seejuures moodustab isa aditiivgeneetilisele mõjule vastav varieeruvus ¼ kogu aditiivgeneetilisest dispersioonist (miks?) ning aditiivgeneetiline dispersioon on hinnatav kujul $\sigma_A^2 = 4\sigma_S^2$.

Olgu meil vaatluse all *a* isa, kellest igaüks on andmetes esindatud täpselt *n* järglasega (tütrega), fenotüübiväärtusi on seega kokku $N = an$. Dispersioonanalüüsi arvutused[7] sellise tasakaalulise, mudelile (2.15) vastava andmestiku korral on koondatud tabelisse 2.2.

ANOVA-hinnangud dispersioonikomponentidele σ_S^2 ja σ_E^2 saadakse, võrdsustades keskruudud nende keskväärtustega:

$$\hat{\sigma}_S^2 = \tfrac{1}{n}\left[\mathrm{MS}(S) - \mathrm{MS}(E)\right] \quad \text{ja} \quad \hat{\sigma}_E^2 = \mathrm{MS}(E).$$

Tabel 2.2. Dispersioonanalüüs tabel mudeli (2.15) korral.

Varieeruvse allikas	Ruutude summa	Vabadus-astmete arv	Dispersiooni hinnang (e keskruut)	Hinnangu keskväärtus
Isa (S_i)	$\mathrm{SS}(S) = \frac{\sum_{i=1}^{a}(\sum_{j=1}^{n} y_{ij})^2}{n} - \frac{(\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij})^2}{an} = \sum_{i=1}^{a} n\bar{y}_{i\cdot}^2 - N\bar{y}_{\cdot\cdot}^2$	$a-1$	$\mathrm{MS}(S) = \mathrm{SS}(S)/a-1$	$n\sigma_S^2 + \sigma_E^2$
Jääk ($E_{ij} = E_{j(i)}$)	$\mathrm{SS}(E) = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{\sum_{i=1}^{a}(\sum_{j=1}^{n} y_{ij})^2}{n} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \sum_{i=1}^{a} n\bar{y}_{i\cdot}^2$	$N-a$	$\mathrm{MS}(E) = \mathrm{SS}(E)/N-a$	σ_E^2

[7] DA läbiviimine on üldjuhul tehniliselt küllaltki keeruline. Paljud arvutused lihtsustuvad, kui eeldada andmete tasakaalulisust (**tasakaalus** (*balanced*) andmed on sellised, kus kõigil faktorite ja nende kombinatsioonide tasemetel on sooritatud võrdne arv mõõtmisi/vaatlusi/vmt) ja mudeli **täielikkust** (st, et mudelis on lisaks nn peamõjudele ja alluvusseoses faktoritele esindatud ka kõikvõimalikud koosmõjud ristseoses faktorite vahel).

Üldreeglid dispersioonanalüüsi arvutuste läbiviimiseks tasakaalus andmete ja täieliku faktorkompleksi korral on järgmised (seejuures kasutame faktorite ja nende tasemete arvude tähistusi nagu *A* ja n_A ning eeldame allutatud faktori tasemete nn taandatud numeratsiooni, st et kui faktor *B* on allutatud faktorile *A*, siis faktori *B* tasemete numeratsioon hakkab iga faktori *A* taseme korral 1-st ja lõpeb n_B-ga).

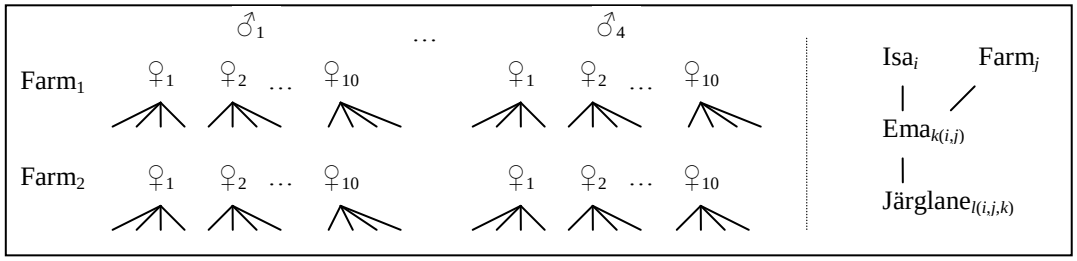
1) Vabadusastmete arvu leidmine.
Vabadusastmete arv faktorile kujul *AB*(*XY*) (faktorite *A* ja *B* koosmõju, mis on allutatud faktorite *X* ja *Y* koosmõjule) on $df_{AB(XY)} = (n_A - 1)(n_B - 1)n_X n_Y$. St, et näiteks $df_A = (n_A - 1)$ ja $df_{AB} = (n_A - 1)(n_B - 1)$.

2) Ruutvormide (*SS*) väärtuste leidmine.
Ruutvormide väärtuste leidmisel kehtib analoogne seaduspära vabadusastmete arvu leidmisega. Tuleb lihtsalt valemeis faktorite tasemete arvud ja nende korrutised, näiteks n_A ja $n_A n_B n_C$ asendada suurustega T_A ja T_{ABC} ja arv 1 suurusega T_μ. Seejuures on oluline avada enne suuruste *n* suurustega *T* asendamist sulud ning arvestada , et $T_A T_B = T_{AB}$. Näiteks $df_{B(A)} = (n_B - 1)n_A$, aga $SS(B(A)) = T_{AB} - T_A$; $df_{AB} = (n_A - 1)(n_B - 1)$, aga $SS(AB) = T_{AB} - T_A - T_B + T_\mu$. Suuruste *T* täpne esitus sõltub mudelis olevate faktorite arvust – näiteks kolmefaktorilise dispersioonanalüüsi mudeli korral (faktorid *A*, *B* ja *C* pluss jääkliige *E*) $T_A = \sum_{i=1}^{n_A}(\sum_{j=1}^{n_B}\sum_{k=1}^{n_C}\sum_{l=1}^{n_E} y_{ijkl})^2 / n_B n_C n_E$, $T_{ABC} = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_C}(\sum_{l=1}^{n_E} y_{ijkl})^2 / n_E$, st et summeeritakse uuritava tunnuse väärtused kõigi suuruse *T* indeksis olevate faktorite tasemete kombinatsioonide korral ja võetakse summa ruutu, summeeritakse kõik saadud ruudud ja jagatakse tulemus läbi *T* indeksis mitte sisaldunud faktorite tasemete arvude korrutisega. Suurus T_μ vastab kõigi uuritava tunnuse väärtuste summa ruudu jagatisele vaatluste arvuga, $T_\mu = (\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_C}\sum_{l=1}^{n_E} y_{ijkl})^2 / n_A n_B n_C n_E$.

3) Dispersiooni hinnangute keskväärtuste E(*MS*) leidmine.
Mingile faktorile *A* vastav dispersiooni hinnangu keskväärtus sisaldab kõigile neile faktoreile vastavaid dispersioonikomponente, mis sisaldavad faktori *A* indeksite komplekti, kusjuures iga dispersioonikomponendi kordaja on nende faktorite tasemete arvude korrutis, mille indeksid ei sisaldu dispersioonikomponendile vastava faktori indeksis. Näiteks eeldades, et eelnevas punktis vaadeldud kolmefaktorilise DA korral faktor *C* on allutatud faktorile *B*, st et DA mudelis on liikmed *A*, *B*, *AB*, *C*(*B*), *AC*(*B*) ja *E*(*ABC*), siis näiteks $\mathrm{E}[MS(AC(B))] = n_E\sigma_{AC(B)}^2 + \sigma_{E(ABC)}^2$ ja $\mathrm{E}(MS(A)) = n_B n_C n_E \sigma_A^2 + n_C n_E \sigma_{AB}^2 + n_E \sigma_{AC(B)}^2 + \sigma_{E(ABC)}^2$. Seejuures jääkdispersioon $\sigma_{E(ABC)}^2 = \sigma_E^2$ sisaldub kõigis E(*MS*) avaldistes ja seda alati kordajaga 1.
Et fikseeritud faktori puhul ei oma dispersioonikomponent mõtet (võrdub definitsiooni kohaselt 0-ga), siis ei ole ka E(*MS*) avaldistes fikseeritud faktori(te)le vastava(te)s liikme(te)s mitte dispersioonikomponendid vaid hoopis sobiva kordajaga suurus, mis väljendab faktori eri tasemetele vastavate mõjude summaarset ruuterinevust faktori keskmisest mõjust. Et aga dispersioonikomponentide hinnangud sellest liidetavast ei sõltu, siis lõpetame ülevaate DA-st siinkohas.

Taoline lihtne mudel võimaldab saada esmase ülevaate uuritava tunnuse geneetilise varieeruvuse suurusest populatsioonis, hinnata isa kui geneetilise faktori mõju olulisust ning analüüsida, kas siis teoreetiliselt või modelleerimiseksperimentide abil, saadavate hinnangute täpsust ja sõltuvust andmete struktuurist.

Järgnevalt on kirjeldatud üht lihtsat modelleerimiseksperimenti, kus püütakse selgitada dispersioonanalüüsist saadavate hinnangute täpsust juhul, kui lisaks isa mõjule sõltub järglase fenotüübiväärtus ka ema aditiivgeneetilist mõjust ja fikseeritud keskkonnaefektist. Kujutame ette (lihtsuse mõttes tasakaalulist) andmestikku, kus nelja isast on ristatud 20 juhusliku emasega, kellest 10 valiti ühest ja 10 teisest farmist. Igast ristamisest võeti 4 järglast, kellel mõõdeti fenotüübiväärtus *y*. Skemaatiliselt on andmete struktuur kujutatud joonisel 2.3.



Joonis 2.3. Mudelile (2.17) aluseks olev andmete struktuur ja seda kajastav nn struktuurigraaf

Ema allub isale, kuna iga ema paariliseks on vaid üks kindel isa. Ema allub ka farmile, kus ta elab. Järglased alluvad farmile oma ema kaudu. Kirjeldatud andmetest tulenev seos järglaste fenotüübiväärtuste ning isa, farmi ja ema vahel on väljendatav lineaarse mudeliga

$$y_{ijkl} = \mu + S_i + F_j + D_{k(ij)} + E_{l(ijk)}, \qquad (2.17)$$

kus indeks $i = 1,...,4$ nummerdab isasid, $j = 1,2$ farme, $k = 1,...,10$ ühele isale vastavaid ühest farmist pärit emasid ja $l = 1,...,4$ sama ema järglasi. Et tüüpiliselt loetakse genotüüp (isa mõju) ja keskkond (farmi mõju) sõltumatuks, on käesolevast mudelist lihtsuse huvides välja jäetud genotüübi ja keskkonna interaktsioon (misläbi mudel ei moodusta enam täielikku faktorkompleksi ja DA läbiviimisel ei saa järgida täpselt eelmise lehe allmärkuses toodud skeemi – isa ja farmi koosmõjule vastavad vabadusastmed ja ruutude summa lisanduvad ema mõjule, kui nii isale kui ka farmile alluvale faktorile).

Kirjeldatud skeemi modelleerib järgmine *SAS*-i programm (joonis 2.4), kus muuhulgas on eeldatud, et uuritav fenotüüp on geneetiliselt määratud 20 geeni poolt, mille igaühe mõju on normaaljaotusega keskväärtusega 0 ja dispersiooniga 1.

Programmis on muuhulgas võimalik muuta isade aretusväärtusi (massiivi „*isaav*" teistsuguse väärtustamisega), geenimõjude varieeruvust (parameetri „*d*" näol) ja täiendavat (geneetilistest interaktsioonidest ja/või teistest keskkonnamõjudest tingitud) jääkvarieeruvust (parameetri „*e*" näol) – praeguse mudeli alusel võime kogu isa ja ema mõjule mittevastava varieeruvuse lugeda tingituks Mendeli valikust.

Vastavalt kasutatud modelleerimisskeemile on aretusväärtuse dispersioon nii isadel, emadel kui ka järglastel 40, sest aretusväärtus on modelleeritud kui 40 sõltumatu geeniväärtuse summa, kus iga geeni väärtus on jaotusega $N(0,1)$. Seega $\sigma_A^2 = 40$. Kuna järglane saab keskmiselt ½ oma isa aretusväärtusest, on isalt saadud aretusväärtuse dispersioon $\sigma_S^2 = \frac{1}{4}\sigma_A^2$. Analoogselt avaldub ka emalt saadud aretusväärtuse dispersioon: $\sigma_D^2 = \frac{1}{4}\sigma_A^2$. Et modelleeritud andmetes puuduvad nii võimalikud geneetilised interaktsioonid kui ka uuritava tunnuse varieeruvust mõjutavad keskkonnafaktorid (parameeter $e = 0$), järeldub valemist (2.17) seos

$$\sigma_y^2 = \sigma_A^2 = \sigma_S^2 + \sigma_D^2 + \sigma_E^2,$$

mille alusel ema ja isa alleelide juhuslikust valikust (Mendeli valikust) tingitud varieeruvus moodustab tervelt poole kogu aditiivgeneetilisest varieeruvusest: $\sigma_E^2 = \frac{1}{2}\sigma_A^2 = 20$.

Toodud programmi abil genereeritud andmed on tasakaalus ja reparametriseeritud (juhuslikud faktorid isa ja ema on nullilise keskväärtusega ja farmide efektide 2 ja -2 summa on null).

```
data av (keep = y isa farm ema laps);
 array isaav[1:4](0 0 0 0); /* isade aretusväärtuste massiiv */
 array misa[1:4, 1:2, 1:20]; /* 4 isa 2 kromosoomi 20-s lookuses paiknevate alleelide mõjude massiiv */
 array mema[1:4, 1:2, 1:10, 1:2, 1:20]; /* 10 ema 2 kromosoomi 20 lookuse mõjude massiiv */
 array mlaps[1:4, 1:2, 1:10, 1:4, 1:2, 1:20];
 /* 4 isa ja 2x10 ema (2 farmi) 4 järglase 2 kromosoomi 20 lookuse mõjude massiiv */
 array mfarm[1:2](2 -2); /* 2 farmi mõjud */
 d=1; my=0; e=0; /* d - alleeliväärtuste dispersioon; my - üldkeskmine; e - jääkvarieeruvus */

 * Vanemate genotüüpide genereerimine;
 do isa=1 to 4; do chrom=1 to 2; do loc=1 to 20;
     misa[isa, chrom, loc] = d*normal(0) + isaav[isa];
 end; end; end;

 do isa=1 to 4; do farm=1 to 2; do ema=1 to 10; do chrom=1 to 2; do loc=1 to 20;
      mema[isa, farm, ema, chrom, loc] = d*normal(0);
 end; end; end; end; end;

 * Järglaste genotüüpide moodustamine (320 järglase 12800 geeni) - mõlema vanema iga lookuse korral
valitakse juhuslikult, kumb alleelidest järglasele pärandub;
 do isa=1 to 4; do farm=1 to 2; do ema=1 to 10; do laps=1 to 4; do loc=1 to 20;
     mlaps[isa, farm, ema, laps, 1, loc] = misa[isa, floor(2*ranuni(0))+1, loc];
     mlaps[isa, farm, ema, laps, 2, loc] = mema[isa, farm, ema, floor(2*ranuni(0))+1, loc];
 end; end; end; end; end;

 * Järglaste genotüüpide arvutamine (320 (=4x10x2x4) fenotüüpi);
 do isa=1 to 4; do farm=1 to 2; do ema=1 to 10; do laps=1 to 4;
  y = my + mfarm[farm] + e*normal(0);
  do loc = 1 to 20;
   y = y + mlaps[isa, farm, ema, laps, 1, loc] + mlaps[isa, farm, ema, laps, 2, loc];
  end; output;
 end; end; end; end; run;
```

Joonis 2.4. Andmete modelleerimisel kasutatud *SAS*-i programm

Nii juhuslike kui ka fikseeritud faktoritega DA on *SAS*-is teostatav statistikaprotseduuriga *GLM* (antud andmete struktuuri puhul sobivad ka protseduurid *VARCOMP* ja *MIXED*), mille programm on toodud joonisel 2.5 ja väljavõte tulemustest joonisel 2.6 (et andmestik on genereeritud juhuslike arvude põhjal, on tulemus igal läbimisel erinev).

```
proc glm data=av;
 class isa farm ema;
 model y = isa farm ema(isa farm);
 random isa ema(isa farm) / test;
run;
```

Joonis 2.5. *SAS*-i protseduuri *GLM* programm

```
                          The GLM Procedure
                       Class Level Information
                Class         Levels    Values
                isa                4    1 2 3 4
                farm               2    1 2
                ema               10    1 2 3 4 5 6 7 8 9 10
                   Number of Observations Used         320
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	8101.33774	102.54858	4.82	<.0001
Error	240	5109.02925	21.28762		
Corrected Total	319	13210.36699			
isa	3	2424.672180	808.224060	14.70	<.0001
farm	1	1553.993483	1553.993483	28.27	<.0001
ema(isa*farm)	75	4122.672075	54.968961	2.58	<.0001

Source	Type III Expected Mean Square
isa	Var(Error) + 4 Var(ema(isa*farm)) + 80 Var(isa)
farm	Var(Error) + 4 Var(ema(isa*farm)) + Q(farm)
ema(isa*farm)	Var(Error) + 4 Var(ema(isa*farm))

Joonis 2.6. Väljavõte *SAS*-i protseduuri *GLM* tulemustest

SAS arvutab muu hulgas faktoritele vastavad keskruudud (*MS*), esitab keskruutude ooteväärtuste E(*MS*) valemid tundmatute dispersiooniparameetrite lineaarkombinatsioonidena (kasutades dispersioonikomponentide hinnanguid mittemõjutava fikseeritud faktori „farm" puhul liiget kujul „*Q*(*farm*)") ja kontrollib hüpoteese faktorite mõjude statistilise olulisuse kohta.

Viimase puhul leitakse keskruudu ooteväärtuste paarid, mis on nullhüpoteesi korral võrdsed ja leitakse *F*-statistik vastavate keskruutude jagatise näol. Näiteks isa mõju statistilise olulisuse kontrollimisel on hüpoteeside paar kujul H_0: $\sigma_s^2 = 0$, H_1: $\sigma_s^2 \neq 0$, mistõttu peaks nullhüpoteesi kehtides olema võrdsed isa mõjule ja ema mõjule vastavad keskruutude ooteväärtused ja *F*-statistiku leiame suhtest F = MS(*S*)/MS(*D*) = 808,22/54,97 = 14,7, mille võrdlemisel *F*-jaotusega vabadusastemete arvudega 3 ja 75 saame tulemuseks, et isa mõju on tugevalt oluline – seega isade aretusväärtus varieerub (ei ole konstantne). Analoogselt arvutades tulevad ka emade aretusväärtuste ja farmide erinevused olulised[8].

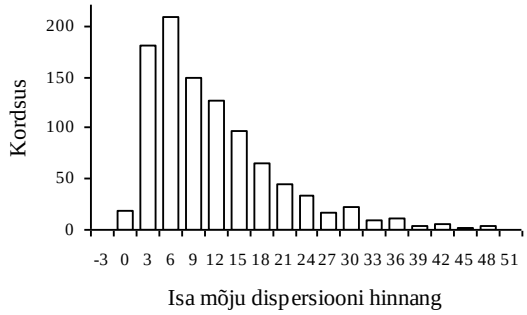
Dispersioonikomponentide hindamiseks võrdsustame keskruudud nende ooteväärtustega ja lahendame saadud lineaarse võrrandisüsteemi σ_S^2, σ_D^2 ja σ_E^2 suhtes:

$$\begin{cases} 808,22 = 80\sigma_S^2 + 4\sigma_D^2 + \sigma_E^2 \\ 54,97 = 4\sigma_D^2 + \sigma_E^2 \\ 21,29 = \sigma_E^2 \end{cases},$$

saades tulemuseks hinnangud $\hat{\sigma}_E^2 = 21,29$, $\hat{\sigma}_D^2 = 8,42$ ja $\hat{\sigma}_S^2 = 9,42$. Aditiivgeneetilise dispersiooni hinnanguks saame isa mõjule vastava dispersioonikomponendi kaudu $\hat{\sigma}_A^2 = 4\hat{\sigma}_S^2 = 37,68$ ja ema mõjule vastava dispersioonikomponendi kaudu $\hat{\sigma}_A^2 = 4\hat{\sigma}_D^2 = 33,68$. Osutub, et mõlemad hinnangud on antud modelleerimiseksperimendi korral pisut väiksemad aretusväärtuse dispersiooni tegelikust väärtusest $\sigma_A^2 = 40$.

Otsustamaks, kuivõrd täpsed on dispersioonikomponentide hinnangud modelleeritud andmete struktuuri korral (see võimaldab hinnata ka analoogse struktuuriga reaalsete andmete analüüsi tulemuste täpsust), kordame kirjeldatud modelleerimist 1000 korda ja hindame iga kord σ_S^2, σ_D^2 ja σ_E^2. Tulemuste statistika on toodud tabelis 2.3 ja dispersioonikomponentide hinnangute empiirilised jaotused joonistel 2.7-2.9.

Modelleerimine kinnitab teoreetilisi arvutusi, sest teoreetilised väärtused jäävad usalduspiiride vahele. Lisaks hakkab silma rida dispersioonikomponentide ANOVA-hinnangute omadusi: hinnangud varieeruvad õige suures ulatuses, kusjuures rohkem varieeruvad need hinnangud, millele vastav vabadusastmete arv on väike; väikese vabadusastmete arvu juures (näiteks σ_S^2 tarvis on vabadusastmeid vaid 3) võivad esineda ka absurdsed negatiivsed hinnangud (taoliste nn illegaalsete hinnangute põhjuseks võib olla ka andmete vähesus ja/või tegeliku parameetri nullilähedane väärtus, lahendusena asendatakse negatiivne hinnang lihtsalt 0-ga); hinnangute jaotused on ebasümmeetrilised, kuid jaotuste keskväärtused kujutavad enesest siiski teoreetiliselt oodatavaid dispersioonikomponentide väärtusi.

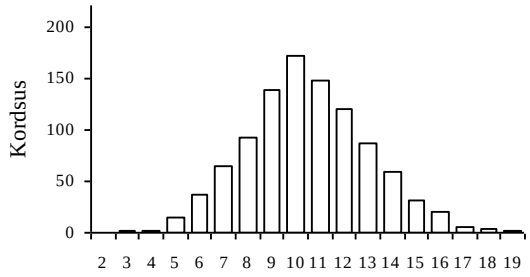


Joonis 2.7. Isa mõju dispersiooni hinnangu empiiriline jaotus

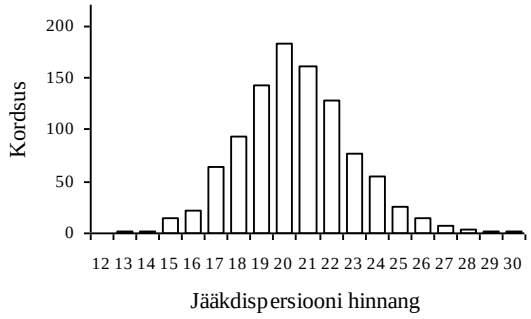
Dispersiooni-komponent	Hinnangute keskmine	95% usalduspiirid	Teoreetiline väärtus
σ_S^2	10,06	(9,51; 10,60)	10
σ_D^2	10,00	(9,85; 10,16)	10
σ_E^2	20,01	(19,86; 20,16)	20

Tabel 2.3. Modelleerimistulemuste kokkuvõte

[8] *SAS* arvutab tegelikult kokku 4 tüüpi keskruute ja nende ooteväärtusi, mis viivad mittetasakaaluliste andmete ja mittetriviaalsete mudelite korral erinevate tulemusteni; samuti on erinevaid variante *F*-statistiku konstrueerimiseks kontrollimaks faktorite mõjude statistilist olulisust (siin kasutatud klassikaliste DA tulemusteni viib lisavaliku „*test*" kasutamine „*random*"-lauses, vaikimisi võrdleb *SAS* kõiki faktoreid juhusliku veaga moodustades *F*-statistiku vaatlusalusele faktorile vastava ja vealiikmele vastava keskruudu jagatisena, http://support.sas.com/onlinedoc/913/docMainpage.jsp).



Joonis 2.8. Ema mõju dispersiooni hinnangu empiiriline jaotus



Joonis 2.9. Jääkdispersiooni hinnangu empiiriline jaotus

2.3.3 Päritavuskoefitsient

Mõõtühikust sõltuva suurusena ei võimalda aretusväärtuse dispersioon siiski vahetult hinnata järglastele päranduva geneetilise potentsiaali (so aretusväärtuse) ja seda peegeldava fenotüübiväärtuse vahelise seose tugevust otsustamaks fenotüübiväärtusel baseeruva selektsiooni mõttekuse üle. Sisuliselt on probleem fenotüübi- ja aretusväärtuse korrelatsioonis. Eeldades fenotüübiväärtuse kujunemist mudeli (2.12) alusel ning kõigi mudelis sisalduvate juhuslike efektide sõltumatust, avaldub huvipakkuv korrelatsioon kujul

$$r_{y_i;A_i} = \frac{\mathrm{cov}(y_i;A_i)}{\sqrt{\mathrm{var}(y_i)\,\mathrm{var}(A_i)}} = \frac{\mathrm{cov}(\mu + A_i + I_i + E_i;A_i)}{\sqrt{\mathrm{var}(\mu + A_i + I_i + E_i)\,\mathrm{var}(A_i)}} = \frac{\sigma_A^2}{\sqrt{(\sigma_A^2+\sigma_I^2+\sigma_E^2)\sigma_A^2}} = \frac{\sqrt{\sigma_A^2}}{\sqrt{\sigma_A^2+\sigma_I^2+\sigma_E^2}}\,. \quad (2.18)$$

Taolise korrelatsioonikordaja ruutu nimetatakse **päritavuskoefitsiendiks** ja tähistatakse h^2:

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2+\sigma_I^2+\sigma_E^2}\,.$$

Seega näitab päritavuskoefitsient, kui suur osa populatsiooni üldisest fenotüübilisest muutlikkusest (mida mõõdab summa $\sigma_A^2+\sigma_I^2+\sigma_E^2=\sigma_y^2$) on tingitud päritavast genotüübilisest muutlikkusest, mistõttu esitatakse päritavuskoefitsient enamasti pisut üldisema seosena aditiivdispersiooni σ_A^2 ja fenotüübidispersiooni σ_y^2 suhtena:

$$h^2 = \sigma_A^2/\sigma_y^2\,.^{9} \quad (2.19)$$

Kuna $0 \le \sigma_A^2 \le \sigma_y^2$, siis $0 \le h^2 \le 1$.

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

Lihtsaim viis kasutada päritavuskoefitsienti indiviidi aretusväärtuse hindamisel on rakendada seda nn kaaluparameetrina, iseloomustamaks indiviidi fenotüübiväärtuse ja populatsiooni keskmise fenotüübiväärtuse erinevuse tingituse määra geenide aditiivsest efektist:

$$\hat{A}_i = h^2(y_i - \mu)\,, \quad (2.20)$$

Taolise valemina jõutakse lihtsast regressioonivõrrandist $A_i = b(y_i - \mu)$, kus regressioonikordaja

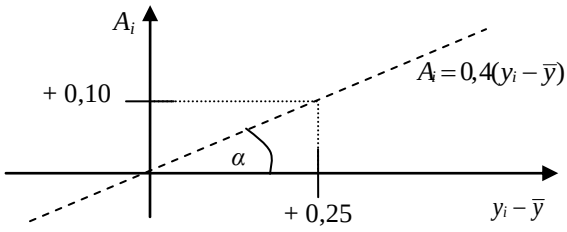
$$b_{A/(y-\mu)} = \frac{\mathrm{cov}(A;y-\mu)}{\mathrm{var}(y-\mu)} = \frac{\mathrm{cov}(A;A)}{\mathrm{var}(y)} = \frac{\sigma_A^2}{\sigma_y^2} = h^2\,.$$

Et regressioonikordaja valem üldkujul rahuldab vähimruutude printsiipi, siis garanteerib seos (2.20) optimaalse (vähima ruutveaga) lineaarse prognoosi indiviidi aretusväärtusele, ja seda nii tegelike populatsiooni parameetrite kui ka nende hinnangute ($\hat{h}^2$, $\hat{\mu}=\bar{y}$) puhul.

[9] Populatsioonigeneetikaalases kirjanduses tähistatakse fenotüübidispersiooni enamasti σ_P^2 (*P – phenotype*), mistõttu ka päritavuskoefitsioent esitatakse kujul $h^2 = \sigma_A^2/\sigma_P^2$.

Näide. Olgu piima rasvaprotsendi päritavuskoefitsient $h^2 = 0{,}40$ ja andku vaatlusalune lehm Lonni keskmisest 0,25 % võrra rammusamat piima. Tema aretusväärtuse hinnanguks on siis +0,10 % (joonis 2.10).

Kuna järglasele pärandub pool ema geenide aditiivsest mõjust, saab Lonni järglane emalt oodatavalt kaasa võime anda keskmisest 0,05 % võrra rammusamat piima.



Joonis 2.10. Regressioonisirge indiviidi *i* aretusväärtuse A_i prognoosimiseks fenotüübiväärtuse hälbe $y_i - \bar{y}$ abil, $h^2 = \tan\alpha = 0{,}40$ korral

Meelde tuleks jätta, et päritavuskoefitsient ei oma mingit absoluutset liigile ja/või tunnusele omast väärtust, vaid on arvutatud konkreetse populatsiooni jaoks konkreetsel ajahetkel. Näiteks tõumaterjali sissetoomine välismaalt suurendab ilmselt geneetilist varieeruvust ja seeläbi ka päritavuskoefitsiendi väärtust. Teisalt mõjutab ka üksnes populatsioonisisese aretuse läbiviimine päritavuskoefitsiendi väärtust, sest sellisel juhul suureneb järjest populatsiooni homogeensus (indiviidid muutuvad geneetiliselt sarnasteks) ning väheneb nii aditiivgeneetiline dispersioon kui ka päritavuskoefitsiendi väärtus. Põllumajandusloomade aretuses on aastakümneid püütud hoida teatud mõttes tasakaalu sise- ja välisaretuse vahel, mistõttu on ka päritavuskoefitsiendid neis populatsioonides suhteliselt stabiilsed ning on võimalik välja töötada aastakümnete pikkusi aretusstrateegiaid (so eeskirju loomade selekteerimiseks saavutamaks mingi ajaga mingi tunnuse keskmise taseme muutust soovitud määral).

Näide. Teada on (Teinberg, R. Põllumajandusloomade geneetika. 1978), et aastatel 1910-1970 suurenes Vändra katselaudas lehmade piima rasvasus 3,3%-lt kuni 4,3%. Võtame piima rasvaprotsendi päritavuseks 0,4 ja veiste põlvkonna pikkuseks 4,5 aastat ning leiame, kui tugev oli valik, st mitme protsendi võrra oli tõuloomade (need, keda kasutati järglaste saamiseks) piim rasvasem populatsiooni keskmisest.

Kokku on piima rasvasus 60 aastaga muutunud $4{,}3\% - 3{,}3\% = 1\%$, milleks on kulunud $60/4{,}5 \approx 13$ põlvkonda.

Eeldame, et tõuloomade valiku kriteerium on nende 60 aasta jooksul olnud sama, st et järglaspõlvkonna vanemaiks on valitud loomad, kelle piima rasvasisaldus on populatsiooni (oma põlvkonna) keskmisest μ (keskmiselt) suuruse Δ võrra suurem, millest vastavalt valemile (2.20) on tõuloomade aretusväärtus $A = h^2\Delta$, millest omakorda järglastele pärandub pool, st et järglaspõlvkonna loomad pärivad ühelt aretusväärtusega $A = h^2\Delta$ vanemalt geneetilise potentsiaali toota vanemate põlvkonna keskmisest $\frac{1}{2}h^2\Delta$ võrra rasvasemat piima. Et pullid piima ei anna, ei saa ka järglaspõlvkonna isasid nende fenotüübiväärtuse alusel valida, seetõttu peame eeldama, et järglaspõlvkonna geneetiline paremus tuleneb üksnes emade valikust ja isad vastavad oma põlvkonna keskmisele (st et üksnes emade aretusväärtused $A_♀ = h^2\Delta$ ja isade aretusväärtused $A_♂ = 0$). Skemaatiliselt on läbiviidud aretustöö kujutatud järgmises tabelis.

Generatsioon	Populatsiooni keskmine fenotüübiväärtus	Valitud emade keskmine		Valitud isade keskmine	
		fenotüübiväärtus	aretusväärtus	fenotüübiväärtus	aretusväärtus
0	$\mu_0 = 3{,}3\%$	$\mu_0 + \Delta$	$h^2\Delta$	μ_0	0
1	$\mu_1 = \mu_0 + \frac{1}{2}h^2\Delta$	$\mu_1 + \Delta$	$h^2\Delta$	μ_1	0
2	$\mu_2 = \mu_1 + \frac{1}{2}h^2\Delta = \mu_0 + 2\cdot\frac{1}{2}h^2\Delta$	$\mu_2 + \Delta$	$h^2\Delta$	μ_2	0
		……...			
13	$\mu_{13} = \mu_{12} + \frac{1}{2}h^2\Delta = \mu_0 + 13\cdot\frac{1}{2}h^2\Delta = 4{,}3\%$				

Valiku kriteeriumi määramiseks saame seega võrrandi $13\cdot\frac{1}{2}h^2\Delta = 1\%$, millest järeldub, et $\Delta = 0{,}38\%$. Seega, 1%-lise piima rasvasisalduse kasvu tarvis pidi 60 aasta jooksul valitud tõuloomade piima rasvasisaldus ületama populatsiooni keskmist keskmiselt 0,38 % võrra.

Ülesanne 13.
Tunnuse päritavuskoefitsient $h^2 = 0{,}7$. Tõuvanemad (mõlemad) valitakse alati populatsiooni keskmisest 0,2 võrra paremad. Mitu põlvkonda on vaja, et näitaja väärtus suureneks 1 ühiku võrra?

Päritavuskoefitsiendi hinnang leitakse dispersioonikomponentide hinnangute suhtena,

$$\hat{h}^2 = \hat{\sigma}_A^2 / \hat{\sigma}_y^2 ,$$

mistõttu on ka hindamismeetodid samad, mis dispersioonikomponentide hindamisel (lihtsamal juhul DA). Sõltuvalt andmete struktuurist ja nende analüüsil kasutatava mudeli kujust, avaldub päritavuskoefitsient erinevalt.

Näiteks valemiga (2.15) esitatud isa mudeli puhul (uuritakse üksnes isalt järglastele pärandunud geenide summaarset mõju) avaldub aditiivgeneetiline dispersioon kujul $\sigma_A^2 = 4\sigma_S^2$ (kus σ_S^2 on isade mõjude dispersioon) ja päritavuskoefitsient kujul

$$h^2 = 4\sigma_S^2 / \sigma_y^2 . \qquad (2.21)$$

Kui uuritavad andmed on tasakaalus (igal isal on võrdne arv järglasi), käib päritavuskoefitsiendi hindamine vastavalt tabelis 2.2 esitatud arvutustele ning päritavuskoefitsient on keskruutude kaudu arvutatav järgmiselt (*n* on järglaste arv isal):

$$\hat{h}^2 = \frac{4[MS(S) - MS(E)]}{MS(S) - (n-1)MS(E)} .$$

Ka keerulisemate mudelite puhul taandub päritavuskoefitsiendi hindamine aditiivgeneetilise dispersiooni hindamisele. Juhul, kui viimast on võimalik mudelist hinnata mitmel moel, on ka päritavuskoefitsiendil mitu võimalikku hinnangut. Näiteks eelmises peatükis modelleerimise aluseks olnud mudelist (2.17) saab aditiivgeneetilist varieeruvust σ_A^2 hinnata 4-l eri viisil – esiteks isade mõjude dispersiooni σ_S^2 kaudu: $\sigma_A^2 = 4\sigma_S^2$, teiseks emade mõjude dispersiooni σ_D^2 kaudu: $\sigma_A^2 = 4\sigma_D^2$, kolmandaks emade ja isade mõjude dispersioonide summa kaudu: $\sigma_A^2 = 2(\sigma_S^2 + \sigma_D^2)$, ja neljandaks, arvestades, et tänu modelleerimisskeemile vastab kogu jääkvarieeruvus järglasele pärandunud geenide juhuslikust valikust tingitud dispersioonile, kujul $\sigma_A^2 = 2\sigma_E^2$. Kasutatud modelleerimisskeemist tuleneb ka, et $\sigma_y^2 = \sigma_A^2$, mistõttu ükskõik millisel viisil leitud aditiivgeneetilise dispersiooni põhjal arvutatud $h^2 = 1$.

Reaalsete andmete korral ei ole võimalik üksteisest eristada Mendeli valiku mõju, geneetiliste interaktsioonide efekti ja juhuslikku keskkonnamõju – need ei pärandu järglastele, mistõttu loetakse kõik need liikmed kuuluvaks mudeli juhusliku vea hulka.

Näide. Vaatame Taani maatõugu sigade andmeid. Uuriti 468 kulti, kellest igaüht ristati 2 emisega. Igast ristamisest mõõdeti 2 samades tingimustes kasvanud isase järglase kehapikkus *y*. Seega on isaga *i* ristatud *j*. ema *k*. järglase kehapikkus esitatav mudeliga

$$y_{ijk} = \mu + S_i + D_{j(i)} + E_{ijk} .$$

Dispersioonanalüüsiga andmeid analüüsides saadi järgmised tulemused.

Mõju	*MS*	E(*MS*)	$\hat{\sigma}^2$
Isa (S_i)	6,03	$4\sigma_S^2 + 2\sigma_D^2 + \sigma_E^2$	$\hat{\sigma}_S^2 = \frac{1}{2}(6{,}03 - 3{,}81) = 0{,}555$
Ema ($D_{j(i)}$)	3,81	$2\sigma_D^2 + \sigma_E^2$	$\hat{\sigma}_D^2 = \frac{1}{2}(3{,}81 - 2{,}87) = 0{,}47$
Jääk ($E_{k(ij)}$)	2,87	σ_E^2	$\hat{\sigma}_E^2 = 2{,}87$
Kokku			$\hat{\sigma}_y^2 = 3{,}895$

Nii emade kui ka isade mõjud on statistiliselt olulised, sest teststatistik emaefektide dispersiooni 0-ga võrdlemiseks on

$$F = MS(D)/MS(E) = 3{,}81/2{,}87 = 1{,}33 \underset{H_0:\sigma_D^2=0}{\sim} F_{468,936}$$

ja isade mõju testimiseks

$$F = MS(S)/MS(D) = 6{,}03/3{,}81 = 1{,}58 \underset{H_0:\sigma_S^2=0}{\sim} F_{467,468}$$

ning mõlema teststatistiku väärtused on kaugelt suuremad kui olulisuse tõenäosusele $p = 0{,}001$ vastavad *F*-jaotuste kriitilised väärtused.

Päritavuskoefitsiendi hinnanguks saame klassikalisel viisil valemist (2.21):
$\hat{h}_{(S)}^2 = 4\hat{\sigma}_S^2 / \hat{\sigma}_y^2 = 4 \cdot 0{,}555/3{,}895 = 0{,}570$.

Et ema mõjule vastava dispersioonikomponendi hinnang $\hat{\sigma}_D^2 < \hat{\sigma}_S^2$, siis ei ole põhjust kahtlustada ema ja keskkonna märgatavat interaktsiooni ja viimase sisaldumist ema mõjuna käsitletavas liikmes (sest muidu oleks emaefektide dispersioon ilmselt üle hinnatud ja suurem, kui isaefektide dispersioon), mistõttu võime päritavuskoefitsienti hinnata ka ema järgi:
$\hat{h}^2_{(D)} = 4\hat{\sigma}_D^2 / \hat{\sigma}_y^2 = 4 \cdot 0,47 / 3,895 = 0,483$,
ning emalt ja isalt pärandunud aditiivgeneetiliste efektide keskmise dispersiooni alusel:
$\hat{h}^2_{(S+D)} = 2(\hat{\sigma}_S^2 + \hat{\sigma}_D^2) / \hat{\sigma}_y^2 = 2(0,555 + 0,47) / 3,895 = 0,526$.

Ülesanne 14.
Kahte isast ristatakse kumbagi 3 emasega ja igast ristamisest saadakse 6 järglast. Tunnuse *y* mõõtmine järglastel andis järgmises tabelis toodud tulemused. **a)** Hinda tunnuse *y* päritavuskoefitsienti h^2. **b)** Kas isa mõju on statistiliselt oluline?

isa	ema	y	isa	ema	y	isa	ema	y	isa	ema	y
1	1	13	1	2	21	2	1	14	2	2	17
1	1	17	1	2	21	2	1	18	2	2	14
1	1	15	1	2	25	2	1	15	2	2	18
1	1	19	1	3	16	2	1	19	2	3	13
1	1	17	1	3	20	2	1	16	2	3	17
1	1	21	1	3	16	2	1	20	2	3	14
1	2	13	1	3	20	2	2	12	2	3	18
1	2	17	1	3	16	2	2	16	2	3	15
1	2	17	1	3	20	2	2	13	2	3	19

2.3.4 Aretusväärtuse hindamine selektsiooniindeksi kujul [10]

Juhul kui kogu uuritav populatsioon paikneb sarnastes keskkonnatingimustes (või on erinevate keskkonnatingimuste mõju täpselt teada ja fenotüübiväärtused selle võrra korrigeeritavad), piisab iga indiviidi aretusväärtuse hindamiseks tema (ja/või tema sugulaste) fenotüübiväärtuste võrdlemisest populatsiooni keskmisega. Kõik need fenotüübil mõõdetud erinevused koondatakse sobivalt valitud kordajatega kaalutuna ühte võrrandisse. Sellist indiviidi aretusväärtuse määramiseks konstrueeritud võrrandit nimetatakse **selektsiooniindeksiks**.

Ühe indiviidi ühe tunnuse aretusväärtust hindava selektsiooniindeksi üldkuju on

$$\hat{A} = I = b_1 X_1 + b_2 X_2 + \ldots + b_m X_m , \qquad (2.22)$$

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).

Maatrikskujul avaldub selektsiooniindeks (2.22) järgmiselt:

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

kus $\mathbf{b} = (b_1\ b_2\ \ldots\ b_m)^{\mathrm{T}}$ on selektsiooniindeksi kaaluparameetrite vektor ja $\mathbf{X} = (X_1\ X_2\ \ldots\ X_m)^{\mathrm{T}}$ on fenotüübil mõõdetud erinevuste vektor [11].

Et selektsiooniindeks on oma kujult mitmene regressioonivõrrand, on tema kordajad b_i avaldatavad seosest

$$b_i = \mathrm{cov}(X_i, A) / \mathrm{var}(X_i) . \qquad (2.23)$$

Sama valem maatrikskujul kirjapanduna:

$$\mathbf{b} = \mathrm{var}(\mathbf{X})^{-1} \mathrm{cov}(\mathbf{X}, A) = \mathbf{P}^{-1}\mathbf{G} , \qquad (2.24)$$

[10] Erinevalt klassikalisest geneetikast, kus selektsiooniindeks näitab mingi genotüübiga isendite populatsioonist kõrvaldamise tõenäosust (peatükk 1.5.2), mõistetakse polügeensete tunnuste analüüsis selle all indiviidi *i* aretusväärtuse A_i hinnangut leituna indiviidi enese ja/või tema sugulaste sobivalt kaalutud fenotüübiväärtuste y_{i_f} summaarse erinevusena populatsiooni keskmisest $\bar{y}$, $A_i = \Sigma_f b_f (y_{i_f} - \bar{y})$. Lihtsaim selektsiooniindeks on (2.14).

[11] Sisuliselt on tegu üldise lineaarse mudeli maatriksesituse (2.5) lahti kirjutamisega ühe elemendi tarvis. Üldjuhul tuletatakse selektsiooniindeksite teoorias valemid maatrikskujul, hindamaks korraga aretusväärtusi paljudele indiviididele (aretusväärtuste vektori näol) või isegi paljude indiviide paljudele (omavahel korreleeruda võivatele) tunnustele (üksikuist aretusväärtuste vektoreist moodustunud maatriksi kujul). Taoline, juba mitmemõõtmelise statistika valdkonda kuuluv mudelite esitus ei mahu aga käesoleva kursuse raamidesse.

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 $m \times 1$-vektor:

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

Järgnevalt vaatame näitena paari konkreetsetel sugulussidemetel baseeruvat selektsiooniindeksit.

Indiviidi aretusväärtuse prognoosimine tema poolõdede fenotüübiväärtuste alusel

Võtame vaatluse alla *n* poolõde, kellel on ühine isa ning erinevad emad. Vaatlusaluse isa iga tulevase tütre aretusväärtus A_* on prognoositav isa praeguste tütarde keskmise fenotüübiväärtuse $\bar{P}_s$ alusel seosest

$$\hat{A}_* = b(\bar{P}_s - \bar{P}), \tag{2.25}$$

kus $b = \text{cov}(A_*, \bar{P}_s - \bar{P}) / \text{var}(\bar{P}_s - \bar{P}) = \text{cov}(A_*, \bar{P}_s) / \text{var}(\bar{P}_s)$, $\bar{P}$ on populatsiooni keskmine fenotüübiväärtus, mis konstantse liidetavana ei oma dispersiooni avaldises mingit rolli.

Isa praeguste järglaste keskmine fenotüübiväärtus avaldub seosena

$$\bar{P}_s = \tfrac{1}{n}\sum_{i=1}^{n}\left(\bar{P} + \tfrac{1}{2}A_s + \tfrac{1}{2}A_{d_i} + E_i\right) = \bar{P} + \tfrac{1}{2}A_s + \tfrac{1}{n}\sum_{i=1}^{n}(\tfrac{1}{2}A_{d_i} + E_i),$$

kus A_s ja A_{d_i} märgivad vastavalt *i*. järglase isa ja ema aretusväärtusi ning E_i kõiki mitte aditiivgeneetilisi mõjusid. Tulevase tütre potentsiaalne fenotüübiväärtus P_* avaldub kujul

$$P_* = \bar{P} + A_* + E_* = \bar{P} + \tfrac{1}{2}A_s + \tfrac{1}{2}A_{d^*} + E_*.$$

Eeldades, et vaatlusalune isa ning olemasolevate ja ka tulevaste tütarde emad ei ole omavahel suguluses, ning et puudub korrellatsioon keskkonna ja genotüübi vahel, jääb tulevase tütre aretusväärtuse ja praeguste tütarde keskmise fenotüübiväärtuse kovariatsioonis alles vaid üks nullist erinev liidetav:

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

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:

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

Siin $\tfrac{1}{4}\sigma_A^2$ on isa aditiivgeneetilisest mõjust tingitud dispersioon ja $\sigma_P^2 - \tfrac{1}{4}\sigma_A^2$ 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 päritavuskoefitsient $h^2 = \sigma_A^2 / \sigma_P^2$, siis järelikult

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

Viimastest võrdustest tulenevalt

$$\text{var}(\bar{P}_s) = \tfrac{1}{4}h^2\sigma_P^2 + \tfrac{1}{n}\left(1 - \tfrac{1}{4}h^2\right)\sigma_P^2,$$

mistõttu avaldub kordaja *b* indeksis (2.25) kujul

$$b = \frac{\text{cov}(A_*, \bar{P}_s)}{\text{var}(\bar{P}_s)} = \frac{\tfrac{1}{4}\sigma_A^2}{\left[\tfrac{1}{4}h^2 + \tfrac{1}{n}(1 - \tfrac{1}{4}h^2)\right]\sigma_P^2} = \frac{nh^2}{4\left[\tfrac{1}{4}nh^2 + (1 - \tfrac{1}{4}h^2)\right]} = \frac{n}{n + [(4 - h^2)/h^2]}. \tag{2.26}$$

Näide. 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 tulevaste tütarde oletatava rasvatoodangu samas karjas.

Vastavalt valemile (2.25) saame Elroi tulevaste tütarde 1. laktatsiooni rasvatoodangu aretusväärtuse hinnanguks:

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

kus

$$b = 25/\left[25 + (4-0,3)/0,3\right] = 0,67,$$

millest

$$\hat{A}_* = 0,67 \times (200 - 230) = -20,1 \text{ kg.}$$

Elroi tulevaste tütarde oletatav 1. laktatsiooni rasvatoodang samas karjas on seega $\hat{P}_* = \bar{P} + \hat{A}_* = 209,9$ kg.

Aretusväärtuse prognoosimine indiviidil enesel ning tema emal ja isal ühekordselt mõõdetud fenotüübiväärtuste alusel

Tähistame indiviidi enese ning tema ema ja isa fenotüübiväärtuste erinevused populatsiooni keskmisest P_o, P_d ja P_s. Aretusväärtuse hindamiseks kasutatav selektsiooniindeksi esitub kujul

$$I = \hat{A}_o = b_1 P_o + b_2 P_d + b_3 P_s . \tag{2.27}$$

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

$$\begin{pmatrix} \sigma^2_{P_o} & \sigma_{P_oP_d} & \sigma_{P_oP_s} \\ \sigma_{P_oP_d} & \sigma^2_{P_d} & \sigma_{P_dP_s} \\ \sigma_{P_oP_s} & \sigma_{P_dP_s} & \sigma^2_{P_s} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} \sigma_{A_oP_o} \\ \sigma_{A_oP_d} \\ \sigma_{A_oP_s} \end{pmatrix}.$$

Eeldades, et kõik fenotüübiväärtused on leitud ühes ja samas populatsioonis, on kõik fenotüübi-dispersioonid võrdsed:

$$\sigma^2_{P_o} = \sigma^2_{P_d} = \sigma^2_{P_s} = \sigma^2_P .$$

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

$$\sigma_{P_oP_d} = \sigma_{P_oP_s} = \tfrac{1}{2}\sigma^2_A = \tfrac{1}{2}h^2\sigma^2_P \quad \text{ja} \quad \sigma_{P_dP_s} = 0 .$$

Analoogsete arutelude tulemusena saame, et

$$\sigma_{A_oP_o} = \sigma^2_A = h^2\sigma^2_P \quad \text{ja} \quad \sigma_{A_oP_d} = \sigma_{A_oP_s} = \tfrac{1}{2}\sigma^2_A = \tfrac{1}{2}h^2\sigma^2_P .$$

Seega on kaaluparameetrid leitavad maatriksvõrdusest

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} \sigma^2_P & \frac{1}{2}h^2\sigma^2_P & \frac{1}{2}h^2\sigma^2_P \\ \frac{1}{2}h^2\sigma^2_P & \sigma^2_P & 0 \\ \frac{1}{2}h^2\sigma^2_P & 0 & \sigma^2_P \end{pmatrix}^{-1} \begin{pmatrix} h^2\sigma^2_P \\ \frac{1}{2}h^2\sigma^2_P \\ \frac{1}{2}h^2\sigma^2_P \end{pmatrix}. \tag{2.28}$$

Näide. Uuritavaks tunnuseks on tallede 100-päeva kehamass päritavusega $h^2 = 0,3$ ja fenotüübidispersiooniga $\sigma^2_P = 84,3$. 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.

Selektsiooniindeksi (2.27) kordajate leidmiseks kirjutame välja maatriksvõrduse (2.28):

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

millest analüüsitava talle 100-päeva kehamassi aretusväärtuseks saame

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