

Tartu Ülikool
Matemaatika-informaatikateaduskond
Matemaatilise statistika instituut

Mirjam Vallas

Juhuslike regressioonikordajatega mudel loomade aretuses

Magistritöö

Juhendajad: lektor Mare Vähi
lektor Tanel Kaart (EPMÜ)

Tartu 2005

Sisukord

Sisukord	2
Sissejuhatus.....	4
1. Juhuslike kordajatega üldise lineaarse segamudeli olemus	6
1.1. Üldine lineaarne segamudel.....	6
1.2. Juhuslike kordajatega üldine lineaarne segamudel.....	10
1.3. Juhuslike kordajatega üldine lineaarne segamudel ja korduvmõõtmiste mudel.....	13
2. Juhuslike kordajatega üldise lineaarse segamudeli rakendused loomade aretuses	16
2.1. Üldine lineaarne segamudel loomade aretuses	16
2.1.1. Isamudel.....	16
2.1.2. Loomamudel	17
2.2. Juhuslike kordajatega üldine lineaarne segamudel loomade aretuses	19
2.2.1. Kasvukõverad	21
2.2.2. Juhuslike regressioonikordajatega mudel loomade elukestusanalüüsis	21
2.2.3. Genotüübi ja keskkonna koosmõju uuringud	22
3. Kontrollpäeva mudel	25
3.1. Kontrollpäeva mudelite tüübid	26
3.1.1. Kahesammuline kontrollpäeva mudel	26
3.1.2. Fikseeritud regressioonikordajatega kontrollpäeva mudel	27
3.1.3. Juhuslike regressioonikordajatega kontrollpäeva mudel.....	29
3.1.4. Mitmemõõtmeline kontrollpäeva mudel.....	32
3.1.5. Kovariatsioonifunktsiooni mudelid	33
3.1.6. Erinevat tüüpi kontrollpäeva mudelite eelised ja puudused	35
3.2. Juhuslike kordajatega funktsiooni valimine	36
4. Rakendusuuring “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” ja juhuslike kordajatega üldise lineaarse segamudeli rakendamine selle raames.....	39
4.1. Rakendusuuring “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” ja juhuslike kordajatega üldise lineaarse segamudeli vajadus selle raames	39
4.2. Juhuslike kordajatega üldise lineaarse segamudeli rakendamine piimatoodangu uurimiseks.....	41
4.2.1. Andmestiku kirjeldus.....	41
4.2.2. Mudelid ja kovariatsioonistruktuuri sobitamine	43

4.2.3. Prognoosid ja parima mudeli valik	47
4.2.4. Pullide ECM-toodangu aretusväärtuste hindamine leitud prognooside abil	52
Kokkuvõte.....	56
Summary.....	58
Kirjandus.....	60
LISA1	63
LISA2	64

Sissejuhatus

Regressioon- ja dispersioonanalüüsi mudelid ja neid hõlmavad üldised lineaarsed mudelid on tänapäeval enim rakendatavad matemaatilise statistika meetodid. Seoses arvutustehnika kiire arenguga on saanud võimalikuks töömahukate hindamisalgoritmide kasutamine, mistõttu analüüsitava mudeli valikul saab enam tähelepanu pöörata reaalsuse täpsemale peegeldamisele. Kuna järjest lihtsamaks on muutunud ka andmete kogumine ja sageli kasutatakse ajas korduvaid vaatlusi, tuleb uuritavat tunnust kirjeldavais mudeleis sageli arvestada korduvate mõõtmistega. Viimase tarvis lisatakse lihtsamal juhul mudelisse korduvatele mõõtmistele vastavate jääkliikmete korreleeritust väljendavad komponendid, keerulisemal juhul püütakse aga võimalikult täpselt modelleerida igal objektil sooritatud korduvate mõõtmiste kulgu. Mudelid, mis taolist uuritava tunnuse korduvalt mõõdetud väärtuste muutust püüavad kajastada, on kasutusel küllaltki laialdaselt, kuid on eri valdkondades või koolkondades esitatud enamasti teatud spetsiifilisel kujul. Näiteks on taimede ja loomade kasvu uurimisel kasutusel üht tüüpi mudelid, mingi keemilise näitaja muutumist sõltuvalt temperatuurist või geograafilisest asukohast kirjeldavad teist tüüpi mudelid, elukestuse uurimisel rakendatakse kolmandat tüüpi mudeleid jne. Loomakasvatusteadus ja geneetika on uuritavate tunnuste ja andmete struktuuri poolest sedavõrd mitmekesised, et neis valdkondades põimuvad omavahel enamus muudes valdkondades rakendatavaist mudeleist.

Vajadus uute mudelite järele tekkis loomakasvatusteadustes ja geneetikas, kust on alguse saanud paljud uued suunad üldiste lineaarsete mudelite teoorias, nagu ikka majanduslikest kaalutlustest lähtuvalt. On ju loomulik, et indiviidi geneetilist potentsiaali saab täpsemalt hinnata juhul, kui selle indiviidi jõudlusnäitajate kohta on teada enam kui vaid üks summaarne number. Ajas korduvalt mõõdetud näitajate kasutamine teeb võimalikuks välja selekteerida need individid, kelle jõudlusomadused muutuvad kõige optimaalsemalt. Samuti saab ajas muutuvana hinnata populatsiooni geneetilisi parameetreid, nagu näiteks päritavust. See omakorda võimaldab selekteerida geneetiliselt paremad individid kõrgema päritavusega vaatlustele tuginedes. Levinud termin taoliste mudelite kohta loomakasvatusteadustes on juhuslike regressioonikordajatega mudel (*random regression models*) või sisule lähedasemalt – juhuslike kordajatega üldine lineaarne segamudel.

Veel üks juhuslike regressioonikordajatega mudelite rakendus on puuduvate väärtuste prognoosimine. Situatsioonis, kus osadel individidel on mõned korduvate mõõtmiste

väärtused puudu, saab need prognoosida, arvestades andmestikust nähtuvat üldist trendi ja lisades sinna konkreetse indiviidi eripära.

Käesoleva magistritöö esimeses osas anname ülevaate juhuslike kordajatega üldise lineaarse segamudeli olemusest. Töö teises osas kirjeldame selle mudeli rakendusi loomade aretuses. Mitmed rakendusvaldkonnad sel alal on praegusel hetkel alles esmaste uuringute tasandil ja juhuslike kordajatega üldine lineaarne segamudel on järjest populaarsust kogumas. Magistritöö kolmas osa räägib lähemalt peamiselt piimakarja jõudlusnäitajate geneetilises hindamises levinud kontrollpäeva mudelist, mida antud hetkel kasutatakse ka Eesti Jõudluskontrolli Keskuses, järjestamiseks tõuloomi ja valimaks geneetiliselt paremaid isendeid järglaspõlvkonna vanemaiks. Anname ülevaate nimetatud mudeli erinevatest tüüpidest, millest üks on juhuslike regressioonikordajatega kontrollpäeva mudel. Tutvustame ka funktsioone, mida kontrollpäeva mudelites kasutatakse, kirjeldamaks uuritava tunnuse ajas kulgemise trajektoori. Töö neljandas osas rakendame erinevaid juhuslike kordajatega üldisi lineaarseid segamudeleid EPMÜ veterinaarmeditsiini ja loomakasvatuse instituudi rakendusuuringu “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” andmete analüüsil. Kasutades varasemates uuringutes esitatud juhuslike kordajatega funktsiooni kujusid, valime välja sobivaima mudeli energia järgi korrigeeritud piimatoodangu kirjeldamiseks laktatsiooni esimesel poolel, prognoosime igapäevased toodangunäitajad kõigile andmestikku kuuluvatele lehmadele ning hindame pullide aretusväärtuste trajektoore tuginedes nende tütarde toodangunäitajate erinevusele karja keskmisest.

Antud magistritöös esinevad loomakasvatusalased mõisted on lahti seletatud töö lisas 1.

1. Juhuslike kordajatega üldise lineaarse segamudeli olemus

1.1. Üldine lineaarne segamudel

Peatükis esitatu põhineb raamatul [1] ning loengumaterjalidel [2] ja [3]. Üldine lineaarne mudel on seos uuritavate tunnuste ja sõltumatute tunnuste ehk faktorite vahel.

Olenevalt uurija eesmärkidest ja analüüsitavatest andmetest omistatakse igale faktorile kindel tüüp – fikseeritud või juhuslik.

Fikseeritud faktoril on vähe tasemeid, iga tase pakub iseseisvat huvi ja on valitud mittejuhuslikult. Andmetes on esindatud kõik huvipakkuvad tasemed. Seetõttu käsitletakse fikseeritud faktorite tasemetele vastavaid efekte kui konstante.

Juhuslikul faktoril on potentsiaalselt väga palju (lõpmatu hulk) tasemeid ja andmetes on neist esindatud valim. Huvi pakub see, kui suur osa uuritava tunnuse koguvarieeruvusest on kirjeldatud antud faktori poolt ehk faktori kõigi võimalike tasemete keskmine mõju. Juhuslike faktorite tasemetele vastavaid efekte käsitletakse kui mingi teoreetilise jaotusega juhuslike suuruste realiseerunud väärtusi. Üldjuhul on selleks teoreetiliseks jaotuseks normaaljaotus.

Fikseeritud efektidega üldine lineaarne mudel on tavalise kordusteta kahefaktorilise dispersioonanalüüsi mudeli näitel esitatav kujul

$$y_{ij} = \mu + a_i + b_j + \varepsilon_{ij}, \quad (1.1)$$

kus a_i , $i = 1, \dots, r$, on esimese faktori i -nda taseme mõju, b_j , $j = 1, \dots, s$, on teise faktori j -nda taseme mõju, y_{ij} on esimese ja teise faktori i -nda ja j -nda taseme kombinatsioonile vastav vaatlus, μ on üldkeskmine ning $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ on juhuslik viga. Ainuke mittekonstantne liige mudeli (1.1) paremal poolel on ε_{ij} .

Üldine lineaarne mudel on maatrikskujul esitatav järgmiselt:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.2)$$

kus juhuslike vigade vektor on normaaljaotusega $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$. Vaatluste vektor on $\mathbf{y}$, maatriks $\mathbf{X}$ on plaanimaatriks ning vektor $\boldsymbol{\beta}$ sisaldab mudeli (1.1) puhul üldkeskmisele μ ja fikseeritud efektidele a_i , b_j vastavaid parameetreid.

Üldine lineaarne segamudel sisaldab nii fikseeritud kui ka juhuslikele faktoritele vastavaid efekte. Lugesdes mudelis (1.1) teise faktori juhuslikuks, saame i -ndale fikseeritud ja j -ndale juhuslikule efektile vastava vaatluse y_{ij} avaldada kujul

$$y_{ij} = \mu + a_i + b_j + \varepsilon_{ij}, \quad (1.3)$$

kus a_i , $i=1, \dots, r$, on fikseeritud faktori i -nda taseme mõju, $b_j \sim N(0, \sigma_b^2)$, $j=1, \dots, s$, on juhusliku faktori j -nda taseme mõju, μ on üldkeskmine ning $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ juhuslik viga. Seega on fikseeritud efektidega mudeli ja juhuslike efekte sisaldava mudeli vaatlusteviisi esitused (1.1) ja (1.3) identsed, erinevus ilmneb efektide jaotustes. Juhuslike efektide $b_j \sim N(0, \sigma_b^2)$ ja $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ kohta eeldatakse traditsiooniliselt, et nad on sõltumatud:

$$\text{cov}(b_j, \varepsilon_{i'j'}) = 0, \quad \forall i, i' \text{ ja } j'.$$

Tavaliselt eeldatakse ka, et b_j erinevate paaride vaheline kovariatsioon on null

$$\text{cov}(b_j, b_{j'}) = 0 \quad \forall j \neq j' \quad (1.4)$$

ning analoogiliselt ka juhuslike vigade puhul

$$\text{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0 \quad \forall i \neq i' \text{ ja } j \neq j'. \quad (1.5)$$

Juhtudel $i = i'$ ja $j = j'$ avaldistes (1.4) ja (1.5) on dispersioonid

$$\text{var}(b_j) = \sigma_b^2 \quad \forall j \text{ ning } \text{var}(\varepsilon_{ij}) = \sigma_\varepsilon^2 \quad \forall i \text{ ja } j. \quad (1.6)$$

Dispersioone σ_b^2 ja σ_ε^2 nimetatakse dispersioonikomponentideks, kuna nad on vaatluse dispersiooni komponendid:

$$\sigma_y^2 = \text{var}(y_{ij}) = \text{var}(\mu + b_j + \varepsilon_{ij}) = \sigma_b^2 + \sigma_\varepsilon^2.$$

Näide 1.1

Olgu meil ühe fikseeritud faktoriga (3 taset, $r = 3$) ja ühe juhusliku faktoriga (kaks taset, $s = 2$) üldine lineaarne segamudel (1.3). Siin $i = 1, 2, 3$ ja $j = 1, 2$. Vaatleme lihtsamat juhtu, kus tegu vaid 6-vaatluselise valimiga (igal faktorite tasemete kombinatsioonil on üks vaatlus). Esitame vastava mudeli maatrikskujul:

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}_{\mathbf{u}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{pmatrix}}_{\boldsymbol{\varepsilon}},$$

kus $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ ning $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$.

Üldine lineaarne segamudel on maatrikskujul järgmine:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (1.7)$$

kus $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ ning $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$.

Üks peamisi erinevusi fikseeritud efektidega üldise lineaarse (1.2) ja üldise lineaarse segamudeli (1.7) maatriksesituses on see, et üldist lineaarset segamudelit esitatakse tavaliselt kahe plaanimaatriksi $\mathbf{X}$ ja $\mathbf{Z}$ abil, kus $\mathbf{X}$ vastab fikseeritud efektidele ning $\mathbf{Z}$ juhuslikele efektidele. Vektor $\boldsymbol{\beta}$ koosneb fikseeritud efektidele vastavatest parameetritest ning vektor $\mathbf{u}$ juhuslikele efektidele vastavatest parameetritest.

Juhuslike efektide kovariatsioonimaatriks avaldub (1.4) ja (1.6) põhjal kujul

$$\mathbf{G} = \sigma_b^2 \mathbf{I}_s = \underbrace{\begin{pmatrix} \sigma_b^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_b^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_b^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \sigma_b^2 \end{pmatrix}}_{s \times s}, \quad (1.8)$$

kus s on juhusliku faktori tasemete arv. Kui juhuslikke faktoreid on rohkem kui üks, on maatriks $\mathbf{G}$ blokkdiagonaalne – igale faktorile vastab tema tasemete arvule vastava dimensiooniga diagonaalmaatriks.

Juhuslike vigade kovariatsioonimaatriks on (1.5) ja (1.6) põhjal

$$\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_n, \quad (1.9)$$

kus n on valimi maht.

Eeldatakse, et maatriksid $\mathbf{G}$ ja $\mathbf{R}$ on mittesingulaarsed ja positiivselt määratud ruutmaatriksid.

Vaatluste vektori $\mathbf{y}$ kovariatsioonimaatriksi $\mathbf{V}$ saab mudelile (1.7) ja selle parameetrite kovariatsioonistruktuurile tuginedes avaldada järgmiselt:

$$\begin{aligned} \mathbf{V} &= \text{var}(\mathbf{y}) = \text{var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) && \{\text{mudelist}\} \\ &= \text{var}(\mathbf{X}\boldsymbol{\beta}) + \text{var}(\mathbf{Z}\mathbf{u}) + \text{var}(\boldsymbol{\varepsilon}) && \{\text{liidetavate sõltumatus}\} \\ &= \text{var}(\mathbf{Z}\mathbf{u}) + \text{var}(\boldsymbol{\varepsilon}) && \{\text{fikseeritud efektide dispersioon on null}\} \\ &= \mathbf{Z} \text{var}(\mathbf{u}) \mathbf{Z}' + \text{var}(\boldsymbol{\varepsilon}) && \{\mathbf{Z} \text{ on konstant}\} \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} && \{\text{mudelist}\} \end{aligned} \quad (1.10)$$

Vaatluste vektori $\mathbf{y}$ keskväärtes avaldub kujul:

$$\begin{aligned} E(\mathbf{y}) &= E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) && \{\text{mudelist}\} \\ &= E(\mathbf{X}\boldsymbol{\beta}) + E(\mathbf{Z}\mathbf{u}) + E(\boldsymbol{\varepsilon}) && \{\text{keskväärtuse aditiivsus}\} \\ &= E(\mathbf{X}\boldsymbol{\beta}) && \{\text{eeldus } E(\mathbf{u}) = E(\boldsymbol{\varepsilon}) = \mathbf{0}\} \\ &= \mathbf{X}\boldsymbol{\beta} && \{\mathbf{X}\boldsymbol{\beta} \text{ on konstant}\} \end{aligned} \quad (1.11)$$

Mudeli (1.7) juhuslike efektide normaaljaotusest tulenevalt ning vaatluste avaldatud keskvaartuse (1.11) ja dispersiooni (1.10) põhjal on vaatluste vektor normaaljaotusega

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{ZGZ}' + \mathbf{R}) \quad (1.12)$$

Näites 1.1 avalduvad kovariatsioonimaatriksid $\mathbf{G}$, $\mathbf{R}$ ja $\mathbf{V}$ vastavalt valemitele (1.8), (1.9) ja (1.10) järgmiselt:

$$\mathbf{G} = \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix},$$

$$\mathbf{V} = \begin{pmatrix} \sigma_b^2 + \sigma_\varepsilon^2 & \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 + \sigma_\varepsilon^2 & \sigma_b^2 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_b^2 + \sigma_\varepsilon^2 & \sigma_b^2 & \sigma_b^2 \\ 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2 + \sigma_\varepsilon^2 & \sigma_b^2 \\ 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_\varepsilon^2 \end{pmatrix}.$$

1.2. Juhuslike kordajatega üldine lineaarne segamudel

Juhuslike regressioonikordajatega mudel tekib lihtsa regressioonimudeli laiendamisel sarnaselt eelmises punktis kirjeldatud fikseeritud efektidega dispersioonanalüüsi mudeli laiendamisele juhuslike efektidega dispersioonanalüüsi mudeliks.

Olgu meil valim ühe subjekti vaatluspaaridest $(x_1, y_1), \dots (x_n, y_n)$, kusjuures huvi pakub tunnuste x ja y vaheline seos. Lihtne regressioonimudel kirjeldamaks argumenttunnuse x mõju uuritavale tunnusele y avaldub kujul:

$$y_j = a + bx_j + \varepsilon_j,$$

kus a ja b on konstantsed hinnatavad regressiooniparameetrid ning $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$ on uuritava tunnuse j -nda väärtuse juhuslik erinevus mudeli poolt prognoositud väärtusest.

Eeldades nüüd, et meil on sellised andmed N subjekti kohta ja korduvate mõõtmiste arv i -ndal subjektil on n_i , siis lisades viimasele mudelile subjektispetsiifilised liikmed, saame regressioonisirged igale subjektile eraldi:

$$\begin{aligned} y_{ij} &= a + bx_{ij} + c_i + d_i x_{ij} + \varepsilon_{ij} \\ &= (a + c_i) + (b + d_i)x_{ij} + \varepsilon_{ij}, \end{aligned} \quad (1.13)$$

$$i = 1, \dots, N, \quad j = 1, \dots, n_i.$$

Siin $c_i \sim N(0, \sigma_c^2)$ iseloomustab subjekti keskmist mõju temal mõõdetud vaatlustulemustele, $d_i \sim N(0, \sigma_d^2)$ on subjektispetsiifiline trend ja $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ väljendab i -ndal subjektil sooritatud j -nda mõõtmise kõrvalekallet mudeliga prognoositust.

Tähistades $a_i = a + c_i$ ja $b_i = b + d_i$, saame mudelile (1.13) kuju

$$y_{ij} = a_i + b_i x_{ij} + \varepsilon_{ij}, \quad (1.14)$$

kus $a_i \sim N(a, \sigma_c^2)$ ja $b_i \sim N(b, \sigma_d^2)$.

Üldisemalt võib mudeleid (1.13) ja (1.14) nimetada ka juhuslike kordajatega üldisteks lineaarseteks segamudeliteks. Taoliste mudelite maatriksesitus subjekti i jaoks on esitatav kujul

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad (1.15)$$

kus $\mathbf{y}_i$ on subjektile i vastavate uuritava tunnuse väärtuste $n_i \times 1$ vektor, $\boldsymbol{\beta}$ on fikseeritud efektide $p \times 1$ vektor (p on diskreetsete fikseeritud faktorite tasemete mõjude ja pidevatele fikseeritud faktoritele vastavate regressiooniparameetrite summaarne arv), $\mathbf{X}_i$ on fikseeritud efektidele vastav $n_i \times p$ plaanimaatriks, $\mathbf{u}_i$ on juhuslike efektide $q \times 1$ vektor (q on diskreetsete juhuslike faktorite tasemete mõjude ja pidevatele juhuslikele faktoritele vastavate regressiooniparameetrite summaarne arv), $\mathbf{Z}_i$ on juhuslikele efektidele vastav $n_i \times q$ plaanimaatriks ja $\boldsymbol{\varepsilon}_i$ on juhuslike vigade $n_i \times 1$ vektor.

Eeldused mudeli (1.15) juhuslikele efektidele on $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{G})$ ja $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$ ning sellest tulenevalt saame analoogiliselt jaotusele (1.12), et $\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i)$.

Kõigi vaatluste

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1' & \mathbf{y}_2' & \dots & \dots & \mathbf{y}_N' \end{pmatrix}'$$

kovariatsioonimaatriks on

$$\begin{aligned} \mathbf{V} = \text{var}(\mathbf{y}) &= \begin{pmatrix} \text{var}(\mathbf{y}_1) & 0 & 0 & \dots & 0 \\ 0 & \text{var}(\mathbf{y}_2) & 0 & \dots & 0 \\ 0 & 0 & \text{var}(\mathbf{y}_3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \text{var}(\mathbf{y}_N) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Z}_1 \mathbf{G} \mathbf{Z}_1' + \mathbf{R}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 \mathbf{G} \mathbf{Z}_2' + \mathbf{R}_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{Z}_3 \mathbf{G} \mathbf{Z}_3' + \mathbf{R}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \mathbf{Z}_N \mathbf{G} \mathbf{Z}_N' + \mathbf{R}_N \end{pmatrix}. \end{aligned} \quad (1.16)$$

Praktikas on kasutusel ka üldiste lineaarsete segamudelite erijuhud, kus ei eeldata juhuslike vigade sõltumatust ning on võimalik sobitada erinevaid kovariatsioonistruktuure, et saavutada parem kooskõla reaalsete andmetega. Sellest tuleb lähemalt juttu ka järgmises punktis 1.3.

Mudelite (1.13) ja (1.14) maatriksesituses kujul (1.15) on uuritava tunnuse väärtuste vektor $\mathbf{y}_i$ ja plaanimaatriksid $\mathbf{X}_i$ ja $\mathbf{Z}_i$ lahti kirjutatavad järgmiselt:

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{in_i} \end{pmatrix} \text{ ja } \mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \dots & \dots \\ 1 & x_{in_i} \end{pmatrix}.$$

Parameetrite vektorid $\boldsymbol{\beta}$ ja $\mathbf{u}_i$ on kujul

$$\boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix} \text{ ja } \mathbf{u}_i = \begin{pmatrix} c_i \\ d_i \end{pmatrix}.$$

Kovariatsioonimaatriksid $\mathbf{G}$ ja $\mathbf{R}_i$ on sarnaselt esitustele (1.8) ja (1.9) diagonaalmaatriksid:

$$\mathbf{G} = \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_d^2 \end{pmatrix} \text{ ja } \mathbf{R}_i = \sigma_\varepsilon^2 \mathbf{I}_{n_i}.$$

Seega on eeldatud, et omavahel sõltumatud on nii juhuslikud efektid c_i ja d_i kui ka mudeli juhuslikud vead.

1.3. Juhuslike kordajatega üldine lineaarne segamudel ja korduvmõõtmiste mudel

Korduvmõõtmiste mudel on üldine lineaarne mudel, vt (1.1) ja (1.2), kus juhuslike vigade kovariatsioonimaatriks pole diagonaalne. See tähendab, et loobume vaatlustevahelise sõltumatuse eeldusest ning loeme subjektil i sooritatud mõõtmised omavahel korreleerituks. Praktikas käsitletakse subjektina enamasti konkreetset indiviidi – inimest, looma vms.

Nagu kirjeldatud punktis 1.2, on ka juhuslike kordajatega üldise lineaarse segamudeli puhul tegu korduvate mõõtmistega. Lisaks juhuslikule veale on ka juhuslikele faktoritele vastavatel

efektidel kovariatsioonistruktuur $\mathbf{G}$. Seetõttu pole uuritava tunnuse y kovariatsioonimaatriks võrdne juhuslike vigade kovariatsioonimaatriksiga nagu korduvmõõtmiste mudeli puhul.

Esineb ka olukord, kus juhuslike kordajatega üldise lineaarse segamudeli ja korduvmõõtmiste mudeli abil kirjeldatud vaatluste kovariatsioonistruktuur on sama. Seda kirjeldab järgnev näide.

Näide 1.2

Korduvmõõtmiste mudelis avaldub subjekti i vaatlustevaheline kovariatsioonimaatriks avaldub kujul

$$\mathbf{V}_i = \underbrace{\begin{pmatrix} \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 & \sigma^2 & \dots & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 & \dots & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 + \sigma_\varepsilon^2 & \dots & \sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \sigma^2 & \dots & \sigma^2 + \sigma_\varepsilon^2 \end{pmatrix}}_{n_i \times n_i},$$

kus σ^2 on vaatlustevaheline kovariatsioon ja n_i on vaatluste arv subjektil i .

Nüüd vaatame juhuslike kordajatega üldist lineaarset segamudelit, kus loeme juhuslikuks vabaliikme ning eeldame, et juhuslikud vead on sõltumatud. Sel juhul on mudelis (1.15) juhuslikele efektidele vastav plaanimaatriks $n_i \times 1$ ühtede vektor:

$$\mathbf{Z}_i = (1 \ 1 \ \dots \ 1)'$$

Kovariatsioonimaatriksi (1.16) põhjal avaldub antud juhul subjekti i vaatlustevaheline kovariatsioonimaatriks kujul

$$\begin{aligned}
\mathbf{V}_i &= \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{n_i \times 1} \underbrace{\sigma^2 \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \end{pmatrix}}_{1 \times n_i} + \underbrace{\begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix}}_{n_i \times n_i} \\
&= \underbrace{\begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \end{pmatrix}}_{n_i \times n_i} + \underbrace{\begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix}}_{n_i \times n_i} \\
&= \underbrace{\begin{pmatrix} \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 + \sigma_\varepsilon^2 & \cdots & \sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 + \sigma_\varepsilon^2 \end{pmatrix}}_{n_i \times n_i}.
\end{aligned}$$

Seega on antud juhul juhuslike kordajatega üldise lineaarse segamudeli ning korduvmõõtmiste mudeli abil kirjeldatud vaatluste kovariatsioonistruktuurid identsed.

2. Juhuslike kordajatega üldise lineaarse segamudeli rakendused loomade aretuses

2.1. Üldine lineaarne segamudel loomade aretuses

2.1.1. Isamudel

Loomade aretuses on eesmärk geneetiliselt parimate loomade välja valimine järglaspõlvkonna vanemaiks. Tüüpilisem näide on isamudel (*sire model*), mille abil püütakse hinnata iga isa geneetilist mõju järglaste toodangunäitajate headusele ja panna isasloomad selle alusel pingeritta. Üldise lineaarse segamudeli mõistes tähendab see, et juhusliku faktori rollis on isa, kelle mõju püütakse hinnata tema järglaste alusel. Enamasti eeldatakse, et isad ei ole omavahel sugulased ning et kõigilt järglastelt on andmestikus vaid üks vaatlus. Seega on mudelile vastavad kovariatsioonimaatriksid diagonaalsel kujul. Mudeli (1.7) tähistustes on juhuslike isamõjude $\mathbf{u}$ kovariatsioonimaatriks

$$\text{var}(\mathbf{u}) = \mathbf{G} = \text{var} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_s \end{pmatrix} = \sigma_b^2 \mathbf{I}_s = \underbrace{\begin{pmatrix} \sigma_b^2 & 0 & \cdots & 0 \\ 0 & \sigma_b^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_b^2 \end{pmatrix}}_{s \times s},$$

juhuslike vigade kovariatsioonimaatriks

$$\text{var}(\boldsymbol{\varepsilon}) = \mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_n$$

ning uuritava tunnuse kovariatsioonimaatriks avaldub kujul

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R} = \sigma_b^2 \mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n = \underbrace{\begin{pmatrix} \sigma_b^2 + \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_b^2 + \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_b^2 + \sigma_\varepsilon^2 \end{pmatrix}}_{n \times n}.$$

2.1.2. Loomamudel

Tänapäeval enim kasutatava mudeli puhul püütakse hinnata kõigi loomade geneetilist potentsiaali. Sellises mudelis, mida nimetatakse loomamudeliks (*animal model*), on juhuslikuks efektiks looma enese mõju ja loomaefektidele vastavas dispersioonimaatriksis püütakse arvesse võtta kõiki loomadevahelisi sugulussidemeid. Viimase tarvis konstrueeritakse eriline aditiivse geneetilise suguluse maatriks $\mathbf{A}$, mille dimensioon on võrdne kõigi andmestikus esindatud loomade ja nende analüüsil kasutatavate eellaste arvuga. Maatriksi $\mathbf{A}$ elementideks a_{ij} on tõenäosused, et indiviidide i ja j genotüübis esinevad geenid on sama päritolu. Peadiagonaalil, kus võrreldakse indiviidi i genotüüpi tema enda omaga, on seega ühed – indiviidi sugulus iseendaga on üks. Erandiks on olukord, kus indiviidi vanemad on omavahel suguluses. Siis lisandub peadiagonaali elemendile a_{ii} vanematevahelist sugulust iseloomustav konstant (suguluskoefitsient). Sel erijuhul saab a_{ii} olla ka suurem kui üks. Seega mudeli (1.7) tähistes pole antud juhul looma mõjule vastavate juhuslike efektide $\mathbf{u}$ kovariatsioonimaatriks enam diagonaalmaatriks:

$$\text{var}(\mathbf{u}) = \mathbf{G} = \sigma_a^2 \mathbf{A},$$

kus σ_a^2 on aditiivgeneetiline dispersioon. Samas eeldatakse siingi, et ühelt loomalt on mudelis üks vaatlus ning juhuslike vigade kovariatsioonimaatriks on sama kujuga nagu isamudelil. Kuna maatriks $\mathbf{G}$ pole diagonaalne, siis on uuritava tunnuse kovariatsioonimaatriks loomamudeli korral

$$\text{var}(\mathbf{y}) = \mathbf{V} = \sigma_a^2 \mathbf{ZAZ}' + \sigma_e^2 \mathbf{I}_n = \begin{pmatrix} a_{11}\sigma_a^2 + \sigma_e^2 & a_{12}\sigma_a^2 + \sigma_e^2 & \cdots & a_{1n}\sigma_a^2 + \sigma_e^2 \\ a_{12}\sigma_a^2 + \sigma_e^2 & a_{22}\sigma_a^2 + \sigma_e^2 & \cdots & a_{2n}\sigma_a^2 + \sigma_e^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}\sigma_a^2 + \sigma_e^2 & a_{2n}\sigma_a^2 + \sigma_e^2 & \cdots & a_{nn}\sigma_a^2 + \sigma_e^2 \end{pmatrix}.$$

Näide 2.1

Isa- ja loomamudelit on kasutatud veiste aretusväärtuse hindamiseks Jõudluskontrolli Keskuses (JKK). Aastatel 1992 - 1996 hinnati mainitud aretusväärtusi isamudeli abil, edaspidi võeti kasutusele loomamudel.

Jõudlustunnustest hinnatakse eraldi piima, piimarasva ja piimavalgu kogus kilogrammides. Hinnatavaid välimikutunnuseid on üle kümne, nende hulgas tõu tüüp, udar, jalad, üldhinne. Mudelid hinnatakse eraldi eesti punast tõugu (EPK) ja eesti holsteini tõugu (EHF) lehmadele [4].

JKK rakendatavad mudelid:

- Isamudel

$$y_{ijkl} = \mu + K_i + g_j + s_{jk} + \varepsilon_{ijkl},$$

kus μ on üldkeskmise, g_j on pullide geneetilise grupi j efekt, s_{jk} pulli k efekt vastavas geneetilisest grupis j , ε_{ijkl} on jääkefekt. Ülejäänud lehma l toodangut mõjutavad efektid (karja pidamise ja söötmise, poegimisaasta mõju jne) sisalduvad nn karjaefektis K_i . Juhuslikena on antud mudelis käsitletud isa efekti ning jääkefekti [5].

- Loomamudel jõudlustunnustele varasematel aastatel:

$$y_{ijkl} = \mu + v_i + p_j + H_k + a_l + \varepsilon_{ijkl}.$$

Fikseeritud mõjudena on mudelis poegimise vanusklass v_i , poegimisvahemiku pikkusklass p_j ja kari-aasta-sesoon koosmõju H_k , kus aasta on poegimisaasta, mis on jagatud perioodideks jaanuar – mai ning juuni – detsember. Lisaks on mudelis ka hinnatava looma l juhuslik aditiivgeneetiline efekt a_l ehk aretusväärtus, juhuslik jääkliige ε_{ijkl} ja üldkeskmise μ [6].

- Loomamudel välimikutunnuste hindamiseks:

$$y_{ijklm} = \mu + F_i + H_j + v_k + p_l + a_m + \varepsilon_{ijklm}.$$

Mudelisse sisalduvad lisaks üldkeskmisele μ ning juhuslikele looma m aditiivgeneetilisele efektile a_m ja jääkefektile ε_{ijklm} järgmised fikseeritud keskkonnaefektid: klassifitseerija-aasta

F_i , kari(karjaklass)-aasta H_j , vanus esimesel poegimisel v_k , laktatsioonipäevad klassifitseerimisel p_l .

Alates 1999. aastast kasutatakse loomamudelit ka sigade aretusväärtuste hindamiseks [4].

2.2. Juhuslike kordajatega üldine lineaarne segamudel loomade aretuses

Saamaks ülevaadet kasvatatavate loomade jõudlusnäitajate muutumisest, sooritatakse mitmesuguseid kontrollmõõtmisi pidevalt. Geneetiliselt potentsiaalilt paremate loomade täpsema selekteerimise huvides on mõttekas püüda arvestada andmete olemasoleva struktuuriga ja mitte hakata olemasolevate mudelite rakendamise tarvis andmeist eelnevalt kokkuvõtteid tegema. Mudeleiks, mis võimaldaksid arvestada ajas korduvalt mõõdetud jõudlusnäitajatega, osutusid juhuslike kordajatega üldised lineaarsed segamudelid. Antud kontekstis võib neid mudeleid nimetada ka juhuslike regressioonikordajatega mudeliteks.

Võttes juhuslike kordajatega üldises lineaarses segamudelis argumenttunnusteks sobiva lineaarse ajafunktsiooni $f(t)$, st aeg on mudelis pideva argumendina, ei pea mõõtmised subjektidel olema tehtud samadel ajahetkedel ja sama palju arv kordi. Viimane tingimus lubab analüüsi kaasata ka tavapärase korduvate mõõtmiste analüüsi mõistes puuduvate väärtustega subjektid. See tõstab analüüside statistilist võimsust ning vähendab saadavate hinnangute nihet, kuna populatsiooni kohta olevaid andmeid on rohkem kui puuduvate väärtusteta analüüsi rakendamisel [3].

Mudeli (1.13) esituses iseloomustavad fikseeritud efektid a ja b vastavalt uuritava tunnuse y populatsiooni üldkeskmist ja faktori keskmist mõju erinevatel ajahetkedel ning juhuslikud efektid c_i ja d_i vastavalt subjekti individuaalset erinevust populatsiooni üldkeskmisest ja faktori keskmisest mõjust vastava tunnuse osas. Tänu subjektispetsiifilistele komponentidele c_i ja d_i saab juhuslike kordajatega üldise lineaarse segamudeli abil hinnata uuritava tunnuse y kulgemise trajektoori igale subjektile eraldi. Vaadeldud mudelist keerukamad mitmefaktorilised mudelid võivad sisaldada ka fikseeritud faktoreid. Fikseeritud faktorite kaasamine lubab praktikas arvesse võtta näiteks keskkonna mõjusid (kari, pidamistingimused jms).

Nimetatud eeliste tõttu varem kasutusel olnud traditsiooniliste mudelite ees on vaadeldava mudeli üks tuntumaid rakendusi loomade aretuses just piimakarja geneetiline hindamine kontrollpäeva toodangu abil. Kontrollpäeva mudelist lähemalt kolmandas peatükis. Antud valdkonnas on aga võimalik leida ka teisi rakendusi, millest annab hea ülevaate Schaeffer [7].

Juhuslike regressioonikordajatega mudeli võib kirja panna kujul

$$y_{ijknt} = F_i + g(t)_j + \sum_{l=0}^{m_1} a_{kl} x_{ijkl} + \sum_{l=0}^{m_2} p_{kl} x_{ijkl} + \varepsilon_{ijknt} ,$$

kus y_{ijknt} on looma k vaatlus n ajahetkel t , mis kuulub fikseeritud faktori tasemele i ja gruppi j ; F_i on ajaskaalast sõltumatu fikseeritud efekt (nt asukoha efekt või kari-kontrollpäeva efekt); $g(t)_j$ on funktsioon(id), mis kirjeldab/kirjeldavad grupi j kõigi loomade keskmist vaatlust; $\sum_{l=0}^{m_1} a_{kl} x_{ijkl}$ ja $\sum_{l=0}^{m_2} p_{kl} x_{ijkl}$ on juhuslike kordajatega regressioonifunktsioonid ajast sõltuvate kovariantidega x_{ijkl} , looma k aditiivgeneetilisele efektile ja püsivale keskkonnaefektile vastavate regressioonikordajatega a_{kl} ja p_{kl} ; ε_{ijknt} on juhuslik viga keskväärtusega null ning erinevate dispersioonidega iga ajahetke t või t funktsiooni jaoks. Funktsioon $g(t)_j$ võib olla nii lineaarne kui ka mittelineaarne ja iseloomustab uuritava tunnuse kulgu ajas. Kui sellise muutuse trajektoori kuju pole teada või on see mittelineaarne, siis võib $g(t)_j$ olla klassifitseerivate tunnuste kogum. Klassifitseerivad tunnused vajavad küll rohkem vabadusastmeid ning nõuavad suurt vaatluste arvu igale tasemele, kuid samas ei sunni nad defineerima konkreetset trajektoori kuju.

Juhuslike regressioonikordajatega funktsioonid on mõeldud hálbe modelleerimiseks tunnuse keskmise trajektoori ümber nagu eelnevalt kirjeldatud mudeli (1.13) jaoks. Argumentideks on neis mudelites soovitatud standardiseeritud ajaga ortogonaalseid polünoome, millest on lähemalt kirjutatud punktis 3.2. Nimetatud punktis on ka näiteid praktikas kasutatud juhuslike kordajatega funktsioonidest.

2.2.1. Kasvukõverad

Loomade kasvu muutust ajas saab modelleerida juhuslike regressioonikordajatega mudeli abil. Üldiselt on kasv kariloomade puhul majanduslikult oluline tunnus. Kasvuga on seotud söömus, söödaväärindus, rasva ladestumine, lihasmassi areng, luude pikkus ja kehaline konditsioon. Kontrollpäeva piimakoguste ja kasvu analüüs on veidi erinev, kuna kasv on ajas kumulatiivne. See oleks analoogiline juhuga, kui individuaalsete kontrollpäeva toodangute asemel vaadata summeeritud igapäevaseid piimakoguseid.

Geneetilistest efektidest peetakse kasvutunnuste puhul oluliseks eelkõige emapoolseid geneetilisi efekte ja seda eelkõige lihakarjades.

Juhuslike regressioonikordajatega mudeli rakendamisel kasvutunnustele on võtmeküsimusteks vajalik indiviidi mõõtmiste arv, millises vanuses need on sooritatud ning milline on vanuse ülempiir, millest edasi loomad tavaliselt kas tapetakse või ei paku edasine väheintensiivne kasv lihtsalt enam huvi. Nende mõõtmiste teostamise maksumus mõjutab samuti seda, kui sageli ja millal kasvu mõõta. Juhuslike kordajatega regressioonimudel võimaldab selles suhtes rohkem vabadust ja üldiselt kaalutakse loomi kõigis vanustes sünnist täisealiseks saamiseni. Osadel loomadel võib olla palju mõõtmisi, samas kui teistel on neid vaid mõned. Fikseeritud kasvukõverad tuleks hinnata sugude kaupa, sünniaastate siseselt, tõugude või tõuristandite piires ning emasloomade erinevate laktatsioonide siseselt. Selles rakendusvaldkonnas on veel palju tööd ära teha.

Peale loomade saab juhuslike regressioonikordajatega mudelit rakendada erinevatele taimedele, näiteks põllusaagile (mis kasvab kiiresti) või puudele (mis kasvavad aeglaselt). Samuti võib seda mudelit kasutada teatavates tingimustes kasvanud bakterite populatsiooni modelleerimiseks. Juhuslike regressioonikordajatega mudeli rakendamine kasvu andmetele on üks kiirelt arenev uurimisvaldkond.

2.2.2. Juhuslike regressioonikordajatega mudel loomade elukestusanalüüsis

Elukestustunnus on selline tunnus, mida on võimalik mõõta iga päev looma eluaja jooksul, kuid andmetes on tavaliselt kirjas vaatlus vaid ajahetkest, mil loom sureb või välja praagitakse. Juhuslike regressioonikordajatega mudeli rakendamist elukestusanalüüsis soovitasid 1999. aastal Veerkamp, Brotherstone ja Meuwissen. Looma individuaalne

“elukestuskõver” on sirgjoon väärtusel 1 looma sünnist kuni karjast eemaldamiseni mingil põhjusel (nt surm, õnnetus). Sellest hetkest alates kuni vaatlusperioodi lõpuni on pideva sirgjoone väärtuseks 0. Paljude loomade “elukestuskõverate” keskmine annab meile populatsiooni keskmise elukestuskõvera oodatavast väljamineku määrast erinevatel ajahetkedel. Vaatlusperioodi alguses (tavaliselt sünn või esimene poegimine) on peaaegu kõik loomad veel elus, mistõttu tunnuse varieeruvus on nullilähedane. Uuritava tunnuse dispersioon kasvab aja jooksul, kuni populatsioonis on alles jäänud 50% loomadest. Vaatlusperioodi lõpus läheneb dispersioon taas nullile, sest selleks ajaks on peaaegu kõik loomad karjast välja langenud.

Juhuslike regressioonikordajatega mudel elukestuses on esitatav järgmisel kujul:

$$y_{ijkmnt} = S(t)_{ij} + C(t)_{ikm} + \sum_{l=0}^{s_1} a_{nl} x_{ijnl} + \sum_{l=0}^{s_2} p_{nl} x_{ijnl} + \varepsilon_{ijkmnt},$$

kus $S(t)_{ij}$ on samal aastal-sesoonil i esmakordselt poeginud ja nüüd ajaperioodil t karjas j olevate lehmade grupi efekt, $C(t)_{ikm}$ on fikseeritud elukestuskõver esmaspoegimisvanusele k ja toodangutasemele m . Liidetavad $\sum_{l=0}^{s_1} a_{nl} x_{ijnl}$ ja $\sum_{l=0}^{s_2} p_{nl} x_{ijnl}$ on vastavalt looma n aditiivgeneetilist efekti ja püsivat keskkonnaefekti modelleerivad juhuslike kordajatega funktsioonid.

See mudel lubab kaasata veel endiselt karja kuuluvaid loomi (puuduvad vaatlused) ja võimaldab lehmade elukestuse geneetilist hindamist. Tulemusena võib geenide summaarsest toimest määratud potentsiaalse elukestuse (ehk elukestuse aretusväärtuse) esitada teatud arvu kuude üleelanute protsendina peale esmapoegimist. Alternatiivselt võiks esitada kuude arvu peale esimest poegimist, mille jooksul pulli järglastest on karjast eemaldatud 50%.

Juhuslike regressioonikordajatega mudelite alased uuringud elukestuses on alles algusjärgus.

2.2.3. Genotüübi ja keskkonna koosmõju uuringud

Genotüübi ja keskkonna koosmõju tähendab, et loomade järjestus geneetilise potentsiaali poolest pole erinevates keskkonnatingimustes ühesugune. Kui neid keskkonnatingimusi saab järjestada (nt temperatuuri-niiskuse indeksi või karja keskmise toodangu järgi), saab

rakendada juhuslike regressioonikordajatega mudelit, kus ajatunnuse asemel on mainitud järjestavad tunnused. Sellisest mudelist on antud juhul kasu, kui erinevaid tingimusi on rohkem kui mõned kategooriad.

Standberg jt kasutasid juhuslike regressioonikordajatega mudelit genotüübi ja keskkonna koosmõjude uurimiseks põhjamaade punasekirjude veisetõul. Sõltumatu muutuja on kari-aasta keskmine kui hälve kolme riigi (Soome, Norra, Rootsi) kõigi aastate keskmisest. Mudel on kujul

$$y_{ij} = f(X_{ij}) + \sum_{l=0}^1 s_{il} X_{ijl} + \varepsilon_{ij},$$

kus y_{ij} on isa i järglase j uuritava tunnuse väärtus, X_{ij} on isa i järglase j kari-aasta keskmise hälve kolme riigi kõigi aastate keskmisest, $f(X_{ij})$ on fikseeritud regressioonikordajatega lineaarfunktsioon kari-aasta keskmisest hälbest X_{ij} , $\sum_{l=0}^1 s_{il} X_{ijl}$ on juhuslike regressioonikordajatega lineaarfunktsioon isa i mõjule argumentiga X_{ij} ja ε_{ij} on juhuslik viga, millele eeldatakse ühist dispersiooni üle vaadeldavate riikide. Uuritavaks tunnuseks y oli 305 laktatsioonipäeva valgu toodang, uulüpsiperioodi pikkus või ajaintervall poegimisest viimase seemenduseni.

Genotüübi ja keskkonna koosmõju uuringutes kasutatud juhuslike regressioonikordajatega mudelid vajavad veel täiendamist. Kaasama peaks ka fikseeritud kõverad (fenotüübilise seose arvestamiseks) ja uurima kõrgemat järku ortogonaalsete polünoomide sobivust. Kuna igalt loomalt on neis uuringutes kaasatud vaid üks vaatlus (ühes kindlas intervallis või keskkonnas), siis võiks loota, et juhuslike regressioonikordajatega mudel võimaldab esile tuua korrelatsiooni või trendi isade geneetilise mõju ja keskkonna vahel.

Juhuslike regressioonikordajatega mudel on potentsiaalselt rakendatav iga kord, kui on tegu uuritavate tunnuste ja aja või vanuse vahelise seosega. Hea mudel loomakasvatustes peaks sisaldama fikseeritud kordajatega kõveraid erinevatele indiviidide gruppidele, juhuslike kordajatega kõveraid looma aditiivgeneetilisele ja püsivale keskkonnaefektile. Kõige sobivam on mudeli argumentidena kasutada ortogonaalseid polünoome. Parima järgu määramiseks neile polünoomidele on vaja läbi viia veel täiendavaid uuringuid.

Lisaks kirjeldatud rakendustele loomade aretuses võib leida juhuslike regressioonikordajatega mudelite kasutust ka piimakarja sigimisuuringutes ja teiste koduloomade toodangunäitajates (nt kanade munatoodang, lammaste iga-aastane villatoodang jne).

3. Kontrollpäeva mudel

Järgnevalt kirjeldame lähemalt piimakarja toodangunäitajate ning geneetiliste parameetrite hindamises levinud kontrollpäeva mudelit (*test-day model*, KPM). Selle mudeli rakendamiseks vajalik andmestik sisaldab uuritavaid näitajaid mõõdetuna lehmadel lüpsiperioodi ehk laktatsiooni (umbes 305 päeva) vältel korduvalt, tavaliselt igakuiselt või kvartaalselt. Enamasti kasutatakse andmeid kolme esimese laktatsiooni kohta. Enne kontrollpäeva mudelit kasutati samade näitajate analüüsiks punktis 2.1 kirjeldatud loomamudelit, mille rakendamiseks summeeriti igalt lehmalt saadud vaatlused üheks laktatsiooni kogutoodanguks.

Võrreldes varem kasutatud loomamudeliga on kontrollpäeva mudelil mitmeid eeliseid. KPM võimaldab laktatsiooni jooksul muutuva keskkonnaefekti arvesse võtmist konkreetsetel kontrollpäeval, lehmasisese varieeruvuse ning ka lehmade toodangunäitajate erineva muutumise kirjeldamist laktatsiooni lõikes. Samas on tänu juhuslike kordajatega kontrollpäeva mudelist hinnatud parameetritele võimalik arvutada geneetilised näitajad, mille abil hinnata piimatoodangu püsivust peale toodangu kõrgpunkti, mis saabub tavaliselt teisel laktatsioonikuul [8, 9]

Juhuslike regressioonikordajatega mudelit tutvustas 1982. aastal C. R. Jr Henderson. Aastal 1994 soovitasid L. R. Schaeffer ja J. C. M. Dekkers seda mudelit kontrollpäeva toodanguandmete analüüsiks piimakarja aretuses. K. Meyer ja W. G. Hill näitasid 1997. aastal kovariatsioonifunktsioonide ja juhuslike regressioonikordajatega mudeli samaväärsuse. Varaseimates juhuslike regressioonikordajatega kontrollpäeva mudelites modelleeriti juhuslike kordajatega üldise lineaarse segamudeli abil üksnes looma aditiivgeneetiline efekt. Alaline keskkonnaefekt modelleeriti korduvate mõõtmiste mudeli abil. Selline mudel andis aga ebareaalset aditiivgeneetilise dispersiooni hinnanguid. Viimase seitsme aasta jooksul on leidnud juhuslike kordajatega üldine lineaarne segamudel laialdast rakendust paljudes uuringutes ja välja on töötatud märksa sobivamaid mudeleid [8]. Kontrollpäeva andmete puhul on seega ülesandeks lehmasisese ning lehmade vahelise varieeruvuse korrektne arvestamine ja samas võimalikult väheste parameetritega mudeli defineerimine [10]. Kontrollpäeva mudeli parameetrite hindamisel pole eesmärk saada

“õigeid“ hinnanguid, vaid jõuda rahuldava kompromissini mudeli keerukuse ja hinnangute korrektsuse vahel [11].

3.1. Kontrollpäeva mudelite tüübid

Loomade aretuses on viimaste aastate jooksul huvi kontrollpäeva andmete analüüsi vastu järjest kasvanud. Järgnevalt on artiklite [9, 12 , 13] põhjal esitatud ülevaate erinevatest levinumatest kontrollpäeva mudelitest, mis on rakendatavad ka muudes valdkondades, kus esinevad sama struktuuriga andmed. Mudelid on püütud esitada võimalikult ühtset tähistust kasutades, et neid saaks võrrelda.

3.1.1. Kahesammuline kontrollpäeva mudel

Kahesammuline KPM (*two-step test-day model*) on mudel, kus kontrollpäeva andmed korrigeeritakse esimesel sammul kõigepealt kontrollpäeva keskkonnaefektidele, tavaliselt fikseeritud efektidega regressioonimudelit kasutades. Järgnevalt kombineeritakse esimesel sammul saadud mudeli jäägid kogu laktatsiooni iseloomustavaks suurusteks (näiteks laktatsiooni piima kogutoodang), millele seejärel teisel sammul rakendatakse mõnda traditsioonilist mudelit (näiteks isa- või loomamudel).

Näide 3.1

Vaatame kahesammulise mudeli rakendamist lihatõugu pullide (edaspidi pullide) kaalu juurdekasvu hindamiseks [14]. Esimesel sammul rakendati iga pulli kontrollpäeva kaaludele lineaarset regressioonimudelit

$$y = at + \varepsilon ,$$

kus y on pulli kaal kontrollpäeval t , ε on juhuslik viga. Regressioonikordaja a iseloomustab pulli kaalu päevast juurdekasvu. Seda kordajat kontrollperioodi pikkusega (140

päeva) korrutades saadi pulli kontrollperioodi juurdekasv. Seejärel kasutati kontrollperioodi juurdekasvu vaatlusena geneetiliseks hindamiseks loomamudel

$$y_{ijklm} = b_i + v_j + g_k + H_l + a_m + \varepsilon_{ijklm},$$

kus y_{ijklm} on juurdekasv tõust i algvanuses j kontrollgrupist k pärit pullil karja ja aasta koosmõjuga l . Juhuslike efektidena on mudelis käsitletud pulli aditiivgeneetilist efekti a_m ning juhuslikku viga ε_{ijklm} .

3.1.2. Fikseeritud regressioonikordajatega kontrollpäeva mudel

Fikseeritud regressioonikordajatega KPM (*test-day model with fixed regressions*) on mudel, kus laktatsioonikõver on modelleeritud fikseeritud efektide abil ja sama lehma kontrolllõpside vaheline seos on esitatud nagu traditsioonilises korduvate mõõtmiste mudelis. Seda mudelit on üldistatud ka mitme uuritava tunnusega juhule (*multiple lactation test day model*), kus kontrollpäeva vaatlusi käsitletakse korduvate mõõtmistena laktatsioonisiselt ning erinevate tunnustena laktatsioonide lõikes [15]. Selline üldine mudel oli varem kasutusel Saksamaal ja andis eeskujul sarnase mudeli juurutamisel Eestis Jõudluskontrolli Keskuses, kus alates 1999. aasta novembrist hakati jõudlustunnuste aretusväärtusi hindama analoogilise kontrollpäeva loomamudeliga (vt näide 3.2). Mitme laktatsiooniga fikseeritud regressioonikordajatega KPM avaldub kujul

$$y_{ijklm} = H_{il} + \sum_{n=1}^{n_b} \beta_{jln} x_{klmn} + a_{kl} + p_{kl} + \varepsilon_{ijklm}, \quad (3.1)$$

kus y_{ijklm} on lehma k laktatsiooni l kontrollpäeva m toodang, H_{il} on ajast sõltumatu fikseeritud karja ja kontrollpäeva koosmõju i laktatsioonil l , a_{kl} on lehma k juhuslik aditiivgeneetiline efekt laktatsioonil l , p_{kl} on lehmaga k seotud püsiv keskkonnaefekt laktatsioonil l ning ε_{ijklm} on juhuslik viga. Laktatsioonikõver modelleeritakse n_b fikseeritud

parameetriga regressioonifunktsiooni $\sum_{n=1}^{n_b} \beta_{jln} x_{klmn}$ kasutades, kus β_{jln} on laktatsiooni l laktatsioonikõvera j regressioonikordaja n ja x_{klmn} on sellele kordajale n vastav laktatsioonikõverat kirjeldava laktatsioonipäeva funktsiooni element.

Erinevad keskmised laktatsioonikõverad hinnatakse fikseeritud efektide klassidele nagu näiteks vanus, sesoon ja regioon. Edasise üldistusena võib leida eraldi regressioonikõverad ka igale kari-aasta koosmõjule, et võtta arvesse karjade vaheline varieeruvus laktatsioonikõvera kujus. Erinevad söötmisstrateegiad võivad tunduvalt mõjutada laktatsioonikõvera kuju. Lisades mainitud fikseeritud efektidele ka juhuslikud efektid, on tegu lihtsa korduvmõõtmiste mudeliga, mis eeldab kogu laktatsiooni vältel konstantseid dispersioone aditiivgeneetilisele ja püsivale keskkonnaefektile. See tähendab võrdseid korrelatsioone mistahes kahe erineva kontrollpäeva vaatluse vahel, sõltumata sellest, kui kaugel üksteisest need vaatlused ajas paiknevad.

Lehma l laktatsiooni sümmeetriline kontrollpäevade vaheline kovariatsioonimaatriks avaldub fikseeritud regressioonikordajatega KPM korral järgmiselt:

$$\mathbf{V}_l = \begin{pmatrix} \sigma^2 & r\sigma^2 & r\sigma^2 & \dots & r\sigma^2 \\ r\sigma^2 & \sigma^2 & r\sigma^2 & \dots & r\sigma^2 \\ r\sigma^2 & r\sigma^2 & \sigma^2 & \dots & r\sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r\sigma^2 & r\sigma^2 & r\sigma^2 & \dots & \sigma^2 \end{pmatrix},$$

kus $\sigma^2 = \sigma_a^2 + \sigma_p^2 + \sigma_\epsilon^2$ on uuritava tunnuse y dispersioon, mis koosneb aditiivgeneetilisest (σ_a^2), püsivale keskkonnaefektile vastavast (σ_p^2) ja juhusliku vea dispersioonikomponendist (σ_ϵ^2). Korduvaid mõõtmisi arvestav komponent r avaldub kujul

$$r = \frac{\sigma_a^2 + \sigma_p^2}{\sigma^2}$$

ning on antud mudeli korral võrdne uuritava tunnuse kahe kontrollpäeva vahelise korrelatsiooniga.

Näide 3.2

Praegusel ajal kasutab JKK mitme uuritava tunnusega fikseeritud regressioonikordajatega kontrollpäeva mudelit jõudlus- ja udara tervise tunnuste geneetilisel hindamisel [4].

- Mitme tunnusega kontrollpäeva loomamudel jõudlustunnuste hindamiseks

Igale hindamises osalevale loomale hinnatakse konkreetse tunnuse (piim, piimarasv, piimavalk) puhul esimese, teise ja kolmanda laktatsiooni toodangu aretusväärtused. Selleks vaadeldakse analüüsiks kasutatud kontrollpäevade vaatlusi kolmel erineval laktatsioonil kui erinevaid tunnuseid.

$$y_{ijklm} = \mu + H_{il} + S_{kl} + \alpha_{kl1} * C + \alpha_{kl2} * C^2 + \alpha_{kl3} * \ln(1/C) + \alpha_{kl4} * [\ln(1/C)]^2 + p_{lj} + a_{lj} + \varepsilon_{ijklm}, \quad (3.2)$$

kus $C = LP/381$.

kus y_{ijklm} on lehma j kontrollpäeva m toodang laktatsioonil l , μ on üldkeskmine. Fikseeritud efektidena on mudelis laktatsiooni kontrollpäeva ja karja-farmi koosmõju H_{il} ning S_{kl} on faktorite poegimisvanus, poegimissesoon, poegimisvahemik, poegimisaasta, isa staatus grupis, laktatsioon tasemete kombinatsioonil k . Juhuslikud on looma j aditiivgeneetiline efekt a_{jl} , püsiv keskkonnaefekt p_{jl} laktatsioonil l ja mitteseletatav jääkefekt ε_{ijklm} .

3.1.3. Juhuslike regressioonikordajatega kontrollpäeva mudel

Juhuslike regressioonikordajatega KPM (random regression test-day model) on fikseeritud regressioonikordajatega KPM edasiarendus, mille korral mudel avaldub kujul

$$y_{ijklm} = H_{il} + \sum_{n=1}^{n_b} \beta_{jln} x_{klmn} + \sum_{n=1}^{n_a} \alpha_{kln} x_{klmn} + \sum_{n=1}^{n_p} \gamma_{kln} x_{klmn} + \varepsilon_{ijklm}, \quad (3.3)$$

kus lisaks fikseeritud regressioonikordajatega kontrollpäeva mudelis (3.1) esinevatele liikmetele on juhuslike regressioonikordajatega laktatsioonipäeva funktsioonid lehma aditiivgeneetilise ning püsiva keskkonnaefekti kirjeldamiseks. Laktatsioonil l lehmale k vastav juhuslik aditiivgeneetiline ja püsiva keskkonnaefekti regressioonikordaja n on vastavalt α_{kln} ja γ_{kln} .

Juhuslike regressioonikordajatega funktsioon on lisatud selleks, et arvesse võtta looma aditiivgeneetilise ja püsiva keskkonnaefekti individuaalset varieeruvust. Fikseeritud regressioonikordajatega funktsioon $\sum_{n=1}^{n_b} \beta_{jln} x_{klmn}$ on antud mudeli puhul kasutatud selleks, et jagada loomad laktatsioonikõvera sarnasuse alusel eraldi gruppidesse (näiteks vanuse, sesooni järgi). Erinevad regressioonifunktsiooni parameetrite arvud n_b , n_a ja n_p näitavad, et erinevates mudeli osades olevad laktatsioonipäeva funktsioonid pole tingimata samad. Enamasti on tegu suhteliselt lihtsate laktatsioonipäeva funktsioonidega, nagu näiteks polünoomid, ortogonaalsed polünoomid (nt Legendre polünoomid) või Wilminki funktsioon. Kasutatakse ka teisi kasutatavaid funktsioone, sellest lähemalt punktis 3.2.

Võrrelduna mudeliga (3.1) on antud juhul tegu erinevate regressioonisirgetega ka aditiivgeneetilise ja püsiva keskkonnaefekti ehk iga individuaalse looma tasandil. Juhuslike regressioonikordajatega KPM lubab aditiivgeneetilise ja püsiva keskkonnaefekti dispersioonidel laktatsiooni vältel muutuda. Kui suur see varieeruvus on, sõltub laktatsiooni trajektoori kirjeldamiseks valitud funktsioonist. Antud mudel võimaldab eristada loomade vahelist ja ajas muutuvat loomade sisest varieeruvust. Aditiivgeneetiline ja püsiva keskkonnaefekti dispersioon laktatsioonipäeval i on $\sigma_{a_i}^2 = x_i \sum_a x_i'$ ja $\sigma_{p_i}^2 = x_i \sum_p x_i'$, aditiivgeneetiline ja püsiva keskkonnaefekti kovariatsioonid laktatsioonipäevade i ja j vahel on $\sigma_{a_{ij}}^2 = x_i \sum_a x_j'$ ja $\sigma_{p_{ij}}^2 = x_i \sum_p x_j'$, kus $\sum_a$ ja $\sum_p$ on vastavalt aditiivgeneetilise efekti a ja püsiva keskkonnaefekti p kovariatsioonimaatriksid.

Lehma j kontrollpäeva sümmeetriline vaatluste vaheline kovariatsioonistruktuur näeb juhuslike regressioonikordajatega KPM korral välja järgmine:

$$\mathbf{V}_l = \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & r_{13}\sigma_1\sigma_3 & \cdots & r_{1n}\sigma_1\sigma_n \\ r_{21}\sigma_1\sigma_2 & \sigma_2^2 & r_{23}\sigma_2\sigma_3 & \cdots & r_{2n}\sigma_2\sigma_n \\ r_{31}\sigma_1\sigma_3 & r_{32}\sigma_2\sigma_3 & \sigma_3^2 & \cdots & r_{3n}\sigma_3\sigma_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1}\sigma_1\sigma_n & r_{n2}\sigma_2\sigma_n & r_{n3}\sigma_3\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}.$$

Siin $\sigma_i^2 = \sigma_{a_i}^2 + \sigma_{p_i}^2 + \sigma_{\varepsilon_i}^2$ on uuritava tunnuse y dispersioon kontrollpäeval i , mis koosneb aditiivgeneetilisest ($\sigma_{a_i}^2$), püsivale keskkonnaefektile vastavast ($\sigma_{p_i}^2$) ja juhusliku vea

dispersioonikomponendist ($\sigma_{\varepsilon_i}^2$) kontrollpäeval i . Uuritava tunnuse korrelatsioon kontrollpäevade i ja j vahel avaldub kujul

$$r_{ij} = \frac{\sigma_{a_{ij}} + \sigma_{p_{ij}}}{\sqrt{\sigma_i^2 \sigma_j^2}}, \quad (3.4)$$

kus $\sigma_{a_{ij}}$ ja $\sigma_{p_{ij}}$ on vastavalt geneetiline ja püsiv keskkonna kovariatsioon kontrollpäevade i ja j vahel.

Näide 3.3 (näite 3.1 jätk)

Lisaks näites 3.1 kirjeldatud kahesammulisele mudelile rakendati samadele andmetele lihatõugu pullide juurdekasvu analüüsiks ka juhuslike regressioonikordajatega kontrollpäeva mudelit

$$y_{ijklmn} = b_i + \sum_{j=0}^3 v_j x_j + \sum_{k=0}^3 g_k x_k + \sum_{l=0}^3 H_l x_l + \sum_{m=0}^3 p_m x_m + \sum_{n=0}^3 a_n x_n + \varepsilon_{ijklmn},$$

kus b_i on tõu fikseeritud efekt, $\sum_{j=0}^3 v_j x_j$ on algvanuse fikseeritud regressioonikordajatega

funktsioon, $\sum_{k=0}^3 g_k x_k$ on kontrollgrupi fikseeritud regressioonikordajatega funktsioon,

$\sum_{l=0}^3 H_l x_l$ on võõrutusaegse karja ja aasta koosmõju fikseeritud regressioonikordajatega

funktsioon, $\sum_{m=0}^3 p_m x_m$ on pideva keskkonnaefekti fikseeritud regressioonikordajatega

funktsioon, $\sum_{n=0}^3 a_n x_n$ on pulli aditiivgeneetilise efekti fikseeritud regressioonikordajatega

funktsioon ning ε_{ijklmn} on juhuslik viga. Kõik regressiooni abil hinnatud mõjud on esitatud kuuppolünoomi kaudu.

3.1.4. Mitmemõõtmeline kontrollpäeva mudel

Mitmemõõtmeline KPM (multiple-trait test-day model) on mudel, kus iga laktatsioonipäeva või laktatsiooni lõikes intervallidena defineeritud laktatsioonipäeva klassidesse kuuluvad vaatlused loetakse eraldi tunnusteks. Lihtsustatult võib mudeli kirja panna eelnevate mudelite tähistusi järgides:

$$y_{ijklm} = H_{il} + \sum_{n=1}^{n_b} \beta_{jln} x_{klmn} + a_{klm} + \varepsilon_{ijklm}.$$

Siin alaindeks m viitab tunnusele, milleks on lehma k laktatsiooni l kontrollpäeva toodang.

Neid tunnuseid analüüsitakse mitmemõõtmelise analüüsi meetoditega. Selle mudeli eelis on, et ei eeldata kontrollpäevadele eelnevalt kindlaksmääratud dispersioonikovariatsioonistruktuuri. Peamine kasu selle lähenemise puhul on, et see ei nõua juhuslike efektide jaoks täpse laktatsioonikõvera kuju määramist.

Puuduseks on aga see, et tuleb defineerida palju tunnuseid, mistõttu on raskendatud fikseeritud efektide ning aditiivgeneetilisele ja püsivale keskkonnaefektile vastavate dispersioonikomponentide hindamine. Kuna ajas muutuvatele kovariatsioonidele dispersioonidele pole mingit struktuuri määratud, võivad parameetrid olla ettearvamatud. Seda ehk parameetrite hindamiseks kasutatavate suhteliselt väikeste andmestike tõttu. Selline käitumine on vastupidine ootustele silutud kõverast laktatsiooni vältel. Wiggans ja Goddard (1997) kasutasid kanoonilist teisendust, et saada suurest hulgast tunnustest nõ “pseudo tunnused”, mis vastavad kõigi mudelis olevate tunnuste kovariatsioonimaatriksi suurimatele omaväärtustele. Sageli on vaja suhteliselt vähe selliseid “pseudo tunnuseid”, et kirjeldada ära peaaegu kogu andmete varieeruvus.

Kui ühe kuu jooksul on tehtud mitu mõõtmist, tuleb antud mudeli puhul kasutada nende mõõtmiste keskmist. Selliste keskmiste dispersioon on aga väiksem kui algsetel kontrollpäeva vaatlustel.

Avaldame lehma l sümmeetrilise kontrollpäeva vaatluste vahelise kovariatsioonistruktuuri mitmemõõtmelise KPM jaoks ühel laktatsioonil. Selleks olgu n kontrollpäeva laktatsiooni vältel kodeeritud tunnustena 1, 2, ..., n :

$$\mathbf{V}_l = \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & r_{13}\sigma_1\sigma_3 & \cdots & r_{1n}\sigma_1\sigma_n \\ r_{21}\sigma_1\sigma_2 & \sigma_2^2 & r_{23}\sigma_2\sigma_3 & \cdots & r_{2n}\sigma_2\sigma_n \\ r_{31}\sigma_1\sigma_3 & r_{32}\sigma_2\sigma_3 & \sigma_3^2 & \cdots & r_{3n}\sigma_3\sigma_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1}\sigma_1\sigma_n & r_{n2}\sigma_2\sigma_n & r_{n3}\sigma_3\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}.$$

Siin $\sigma_i^2 = \sigma_{a_i}^2 + \sigma_{p_i}^2 + \sigma_{\varepsilon_i}^2$ on uuritava tunnuse dispersioon kontrollpäeval i , mille komponendid on samad nagu kirjeldatud juhuslike regressioonikordajatega KPM puhul. Uuritava tunnuse korrelatsioon kontrollpäevade i ja j vahel avaldub kujul

$$r_{ij} = \frac{\sigma_{a_{ij}} + \sigma_{p_{ij}} + \sigma_{\varepsilon_{ij}}}{\sqrt{\sigma_i^2 \sigma_j^2}},$$

kus lisaks avaldisega (3.4) sarnastele komponentidele on kontrollpäevade i ja j vaatluste vaheline jääkkovariatsioon $\sigma_{\varepsilon_{ij}}$.

Mitmemõõtmelise KPM kovariatsioonistruktuur näeb välja samasugune nagu juhuslike regressioonikordajatega KPM korral. Samas nõuab mitmemõõtmelise KPM puhul kasutatav meetod kõigi dispersiooni-kovariatsioonikomponentide otsest hindamist, juhuslike regressioonikordajatega KPM korral on need komponendid aga laktatsioonipäeva funktsioonid ning laktatsioonikõverat kirjeldava funktsiooni parameetrite dispersioonide-kovariatsioonide funktsioonid. Kui vaadata mitmemõõtmelises kontrollpäeva mudelis tunnustena erinevaid laktatsioonikuid (analüüsitav periood 5 – 305 laktatsioonipäeva), on dispersioonidel ja kovariatsioonidel σ_a^2 , $\sigma_{a_{ij}}$, σ_p^2 , $\sigma_{p_{ij}}$, σ_ε^2 ja $\sigma_{\varepsilon_{ij}}$ 10 erinevat väärtust, samas kui juhuslike regressioonikordajatega kontrollpäeva mudelis on neil komponentidel 300 võimalikku väärtust vastavalt vaatlusaluse perioodi erinevate laktatsioonipäevade arvule.

3.1.5. Kovariatsioonifunktsiooni mudelid

Kovariatsioonifunktsioon (*covariance function*) on funktsioon, mis kirjeldab erinevatel laktatsioonipäevadel mõõdetud vaatluste/tunnuste vahelisi dispersioone ja kovariatsioone.

See funktsioon annab meile kovariatsiooni mistahes kahe teatud ajahetkedel mõõdetud vaatluse vahel ajafunktsioonide ja kovariatsiooniparameetrite avaldisena. Kovariatsioonifunktsioonide kirjeldamiseks sobiv funktsioonide pere on ortogonaalsed polünoomid, millest laialdast kasutust on leidnud Legendre polünoomid. Ortogonaalsed polünoomid on sobivad modelleerimaks nii geneetiliste efektidele kui ka keskkonna efektidele ning fenotüübiväärtustele vastavate kovariatsiooniparameetrite muutumist ajas. Olgu $\phi_j(a_i)$ ortogonaalne j -järku polünoomi ajahetkele i vastav väärtus. Kovariatsioonifunktsiooni $F(a_l, a_m)$, mis näitab ajahetkedele l ja m vastavate vaatluste vahelist kovariatsiooni, saame avaldada kujul

$$F(a_l, a_m) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \phi_i(a_l) \phi_j(a_m) c_{ij}, \quad (3.5)$$

kus k on polünoomi järk, c_{ij} on kovariatsioonifunktsiooni kordaja (enamasti mingi mudelile vastava kovariatsioonimaatriksi parameeter) ja a_m on polünoomide määramispiirkonda standardiseeritud ajahetk m .

Juhul, kui kovariatsiooniparameetrite ajas kulgemist modelleeritakse sama funktsiooni abil nagu ka uuritava tunnuse kulgu, on kovariatsioonifunktsiooni mudel identne tavalise juhuslike regressioonikordajatega mudeliga.

Olgu meil lihtne regressioonimudel

$$y_l = \sum_i \beta_i x_{il} + \varepsilon_l, \quad (3.6)$$

kus y_l tähistab vaatlust indiviidil l , β_i on regressioonikordaja, x_{il} on argumenttunnus ja ε_l jääkliige. Eeldame, et y_l on mõõdetud vanuses a_l , mis on standardiseeritud intervalli $[-1;1]$. Olgu regressioonimudeli (3.6) argument i võrdne Legendre polünoomiga järku i , st $x_{il} = \phi_i(a_l)$, kus $i = 0, \dots, k-1$. Siis saame

$$y_l = \sum_{i=0}^{k-1} \beta_i \phi_i(a_l) + \varepsilon_l$$

ning juhuslike regressioonikordajate jaoks

$$Cov(y_l, y_m) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \phi_i(a_l) \phi_j(a_m) Cov(\beta_i, \beta_j) + Cov(\varepsilon_l, \varepsilon_m). \quad (3.7)$$

Vaadates võrduse (3.7) paremat poolt ilma jääkide kovariatsioonita, kirjeldab see avaldis kovariatsioonifunktsiooni (3.5), kus $Cov(\beta_i, \beta_j)$ on samaväärne kordajaga c_{ij} [16].

Kovariatsioonifunktsioone saab kasutada ka erinevate tunnuste (näiteks piim, piimarasv, piimavalk) vaheliste kovariatsioonide kirjeldamiseks. Eriti mitmemõõtmelisel juhul on see eeliseks, sest seoste modelleerimiseks piisab märksa vähemate parameetrite hindamisest. Sellist taandatud dimensiooniga mudelit kasutatakse näiteks Soomes piimakarja riiklikus hindamissüsteemis.

3.1.6. Erinevat tüüpi kontrollpäeva mudelite eelised ja puudused

Nagu näha, saab kontrollpäeva andmetele rakendada erinevaid mudeleid ning käesolevaks hetkeks pole ühte parimat välja selgitatud. Parima mudeli olemasolu on ka üsna ebatõenäoline, kuna konkreetsed tingimused võivad määrata mudelis vajalikud efektid.

Fikseeritud regressioonikordajatega KPM (FKPM) eeldab indiviididele kogu laktatsiooni vältel konstantset aditiivgeneetilist efekti ega luba laktatsioonikõverate kujus erinevusi loomade tasandil, vaid ainult sarnaste loomade gruppide vahel. Ühelt poolt on selle eelduse tulemuseks ökonoomne parametrisatsioon, teisalt ei näi FKPM kontrollpäevade informatsiooni hästi kasutavana ning ei peegelda hästi mõõdetud väärtusi. Nii mitmemõõtmeline KPM (MKPM) kui ka juhuslike regressioonikordajatega KPM (JKPM) hindavad laktatsioonikõverad igale indiviidile eraldi. MKPM mitmemõõtmeline lähenemine nõuab suurt hulka erinevaid parameetreid, mille hindamine on arvutuslikult väga mahukas (suurte andmestike puhul lausa võimatu). Sellega võivad kaasneda suured standardvead. Lisaks sellele pole mudelil võimalik osa informatsiooni (ühe kuu jooksul tehtud mitmetest mõõtmistest) ära kasutada. Tänu kovariatsioonifunktsioonide rakendamisele juhuslike regressioonikordajatega kontrollpäeva mudelites saab leida dispersiooni ja kovariatsiooni kirjelduse igale laktatsioonipäevale ning seega lubab fenotüübilise informatsiooni optimaalset kasutamist ja järelikult ka korrektsemat parameetrite hindamist.

3.2. Juhuslike kordajatega funktsiooni valimine

Sobivat kuju piimakarja jõudlusnäitajaid iseloomustavatele laktatsioonikõveratele on otsitud juba pikemat aega enne juhuslike kordajatega funktsioonide kasutusele võtmist [17]. Esialgsete, lehmadele vähemalt teatava alagrupi piires ühiste, laktatsioonikõverate asemel hakati 1990ndate aastate esimesel poolel hindama neid kõveraid ka igale lehmale eraldi. Schaeffer ja Dekkers [18] pakkusid välja mudeli, mis erinevalt varasematest sisaldas juhuslikke regressioonikordajaid kontrollpäeva funktsioonile laktatsioonikõvera kuju kirjeldamiseks ning võimaldas seega igale lehmale leida individuaalse laktatsioonikõvera. Mudelis kasutasid artikli autorid juhuslike kordajatega α_0 ja α_1 funktsiooni

$$f(LP) = \alpha_0 * LP + \alpha_1 * \ln(305 / LP),$$

kus LP on laktatsioonipäev.

Juhuslike regressioonikordajatega mudelit, milles erinevalt mudelist (3.3) ei rakendatud regressioonifunktsiooni juhuslikule keskkonnaefektile, kasutasid 1997. aastal Jamrozik ja Schaeffer [19]. Looma aditiivgeneetilist efekti antud mudelis kirjeldas järgmine funktsioon:

$$f(LP) = \alpha_0 + \alpha_1 * C + \alpha_2 * C^2 + \alpha_3 * \ln(1 / C) + \alpha_4 * [\ln(1 / C)]^2, \quad (3.8)$$

kus $C = LP / 305$.

Antud funktsiooni on kasutatud artiklites korduvalt ka võrdluses erinevate funktsioonidega fikseeritud või juhuslike regressioonikordajatega mudeli osa valikuks [20, 21]. Sarnaselt funktsiooniga (3.8) on palju rakendust leidnud ka Wilminki funktsioon

$$f(LP) = \alpha_0 + \alpha_1 * LP + \alpha_2 * \exp(-0.05 * LP). \quad (3.9)$$

Seda funktsiooni kasutati kuni 2003. aasta maikuuni Kanadas, kus ühena esimestest rakendati kontrollpäeva mudelit pullide geneetilisel hindamisel [22].

Viimastel aastatel on paljud uurijad (kaasaarvatud Kanadas [23]) loobunud laktatsioonikõvera kuju kirjeldamisest ülaltoodud ajafunktsiooni abil, kuna aditiivgeneetilisele ja püsivale keskkonnaefektile hinnatavate kovariatsioonimaatriksite

komponendid on sel juhul omavahel tugevasti korreleeritud ning see võib raskendada hindamisprotsessi. Aja muutujaga ortogonaalsetel polünoomidel on komponentide vaheline korrelatsioon palju väiksem ning nende abil saadud kovariatsioonimaatriksite hinnangud kalduvad olema palju robustsemad üle erinevate andmestike. Erinevat tüüpi ortogonaalsete polünoomide (näiteks Legendre versus Tšebõšovi polünoomid) arvutusomadustes võib olla erinevusi. Legendre polünoomid on ühed lihtsamad, mida kasutada [3, 7]. Viimastel aastatel ongi need polünoomid leidnud laialdast kasutust laktatsioonikõvera modelleerimisel [24, 25].

Normeeritud Legendre polünoomid on antud valemiga

$$P'_j(t) = \frac{1}{2^j} \sqrt{\frac{2j+1}{2}} \sum_{m=0}^{[j/2]} (-1)^m \frac{(2j-2m)!}{m!(j-m)(j-2m)!} t^{j-2m},$$

kus $[j/2]$ on suurim täisarv, nii et $[j/2] \leq j/2$.

Esimesed kolm polünoomi avalduvad valemitega:

$$P'_0(t) = 1/\sqrt{2},$$

$$P'_1(t) = \sqrt{3/2} * t,$$

$$P'_2(t) = \sqrt{45/8} * t^2 - \sqrt{5/8}.$$

Legendre polünoomid on defineeritud lõigul $[-1;1]$. Seetõttu tuleb nende rakendamiseks aeg standardiseerida:

$$t = \frac{2 * (LP - LP_{\min})}{(LP_{\max} - LP_{\min})} - 1,$$

kus $LP_{\min}$ on varaseim ajahetk ning $LP_{\max}$ on hiliseim andmetes esinev ajahetk.

Lidauer ja Mäntysaari [26] kasutasid aga Legendre polünoome normeerimata kujul:

$$P_j(t) = \frac{1}{2^j} \sum_{m=0}^{[j/2]} (-1)^m \frac{(2j-2m)!}{m!(j-m)(j-2m)!} t^{j-2m}.$$

Nad leidsid laktatsioonikõvera kuju kirjeldamiseks kompromissi Legendre polünoomide ning Wilminki funktsiooni vahel:

$$f(t) = \alpha_0 * P_0(t) + \alpha_1 * P_1(t) + \alpha_2 * P_2(t) + \alpha_3 * \exp(-0.05 * t) . \quad (3.10)$$

Antud mudelit on kasutanud samuti mitmed uurijad sobivate loomaspetsiifiliste laktatsioonikõverate väljaselgitamiseks võrdluses Wilminki ja erinevat järku Legendre polünoomidega [8, 27].

4. Rakendusuuring “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” ja juhuslike kordajatega üldise lineaarse segamudeli rakendamine selle raames

4.1. Rakendusuuring “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” ja juhuslike kordajatega üldise lineaarse segamudeli vajadus selle raames

EPMÜ Loomakasvatusteaduste instituudi poolt 2001. aastal Põlula katselaudas alustatud rakendusuuring “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” on maailmas suhteliselt ainulaadne kompleksuuring, kus erinevate erialade spetsialistid viivad üheaegselt samadel loomadel läbi paljude erinevate elutalitlusega seotud näitajate uuringuid. Veterinaarid ja sigimist uurivad teadlased fikseerivad kõik seemenduste ja poegimistega seonduva ning määravad regulaarselt erinevate hormoonide taset veres. Söötmise ja ainevahetuse uurijad fikseerivad lehmade poolt söödud söödakogused ning määravad sööda koostise alusel sellest saadava energia (sõltuvalt sööda koostisest jaguneb söödast saadav energia veel paljudeks alaliikideks, mis ka ükshaaval kirja pannakse), samuti määratakse vereproovidest erinevate metaboliitide sisaldus, mille abil püütakse samuti iseloomustada looma organismi energiaga varustatust. Piimajõudluse uurijad koguvad kokku andmed lehmade piimatoodangu ja selle konsistentsi kohta, piimakvaliteedi uurimise labor aga määrab piimaproovidest piima laapumisomadused (viimastest sõltub näiteks piimast saadava juustu kvaliteet). Lisaks kaalutakse lehmi regulaarselt ning hinnatakse nende toitumust visuaalselt. Kokkuvõttes on näiteks söödast saadud energia ning piimatootmiseks ja muuks elutegevuseks kulutatud energia vahena leitav energiabilanss (sõltub lehma kaalust), mis on peamine lehmade optimaalse söötmise uurimisel kasutatav näitaja. Viimast on omakorda võimalik kõrvutada nii piimakvaliteedinäitajatega kui ka sigimist või ainevahetust iseloomustavate näitajatega. Võttes kõigele eelnevale juurde veel ka lehmade põlvnemise ja geneetilised uuringud, on võimalik uurida sigimise, ainevahetuse ja jõudluse geneetilist seost ning determineeritust. Peale selle on tänapäevases majandusruumis haruldasena fikseeritud ja tehtud uurijaile kättesaadavaks kõiksugu kulude ja tuludega seotud numbrid, mis

võimaldab teha kokkuvõtteid ning teostada uuringuid piimakarja pidamise tasuvusest Eesti Vabariigis.

Komplektsete, erinevaid valdkondi hõlmavate uuringute läbiviimine on seni olnud siiski raskendatud. Põhjus peitub selles, et inim- ja finantsressursside piiratuse tõttu ei ole olnud mõeldav uuritavate näitajate mõõtmine ja fikseerimine igapäevaselt. Kõikvõimalikud näitajad on määratud vaid kord või paar kuu jooksul. Andmete struktuuri teeb veel keerulisemaks see, et teatud spetsiifilised proovid võetakse kõigilt lehmadel teatud kindlal kuupäeval (millal vastava eriala spetsialist lauta külastab). Edasise uuringu ajaline telg aga ei järgi mitte kalendrikuupäevi, vaid lehmade sigimistsükli, kus kokkuleppeliselt on esimeseks päevaks lehma poegimispäev, mis on aga erinevatel lehmadel kuupäeva osas erinev. Peale kõige muu on Põlula katselaudas ka lehmade karjast väljalangevus ja prakeerimine suhteliselt suur. Lisaks tuleb aegajalt ette piima- või vereproovi riknemisi või mõõtevigu. See kõik toob aga kaasa suure hulga puuduvate vaatlustega andmestiku.

Selleks, et teostada komplektseid analüüse erinevatel ajahetkedel fikseeritud tunnustega (näiteks piima laapumisomaduste ja verenäitajate või energiabilansi vahelise võimaliku seose uurimine), on vajalik nende tunnuste väärtuste prognoosimine samadeks ajahetkedeks. Arvestades nii konkreetse tunnuse üldist ajalist dünaamikat uuritavas karjas, teatud täiendavate faktorite (poegimisseson, vanus, põlvnemine) mõju kui ka looma enda omapära (väljendatuna temal mõõdetud tunnuse väärtustena), on selliseks individidele puuduvate väärtuste prognoosimiseks sobiv kasutada punktis 1.2. kirjeldatud juhuslike regressioonikordajatega dispersioonanalüüsi segamudelit.

Tulemusena kirjeldab iga looma iga tunnust ajas kulgev kõver. Erinevate tunnuste vaheliste potentsiaalsete seoste uurimiseks ei võrrelda omavahel enamasti mitte kõveraid (neid on hea kasutada tulemuste illustreerimiseks), vaid loetakse iga looma igale tunnusele vastavalt kõveralt välja teatud iseloomustavad näitajad (näiteks maksimum või miinimum, selle saavutamise hetk, langusperioodi pikkus jne). Selliste konkreetse tunnuse käitumist iseloomustavate loomaspetsiifiliste tunnuste alusel on viimaks võimalik läbi viia analüüsid, et kontrollida hüpoteese erinevate näitajate omavahelise korreleerituse kohta.

4.2. Juhuslike kordajatega üldise lineaarse segamudeli rakendamine piimatoodangu uurimiseks

Eesmärk on antud juhul leida juhuslike kordajatega üldine lineaarne segamudel, mille abil võimalikult täpselt hinnata lehmadele individuaalsed energia järgi korrigeeritud piimatoodangu (ECM) laktatsioonikõverad. ECM on üks peamisi lehmade söömust ja piimatoodangut komplekselt kajastavaid näitajaid. Et piimas peituvat energia kogus sõltub eelkõige piima rasva- ja valgusialdusest, arvutatakse ka ECM baseeruvana piima kogutoodangul ning rasva ja valgu osal sellest. Järgnevais uuringuis on ECM arvutatud Soomes kasutusel oleva valemi järgi:

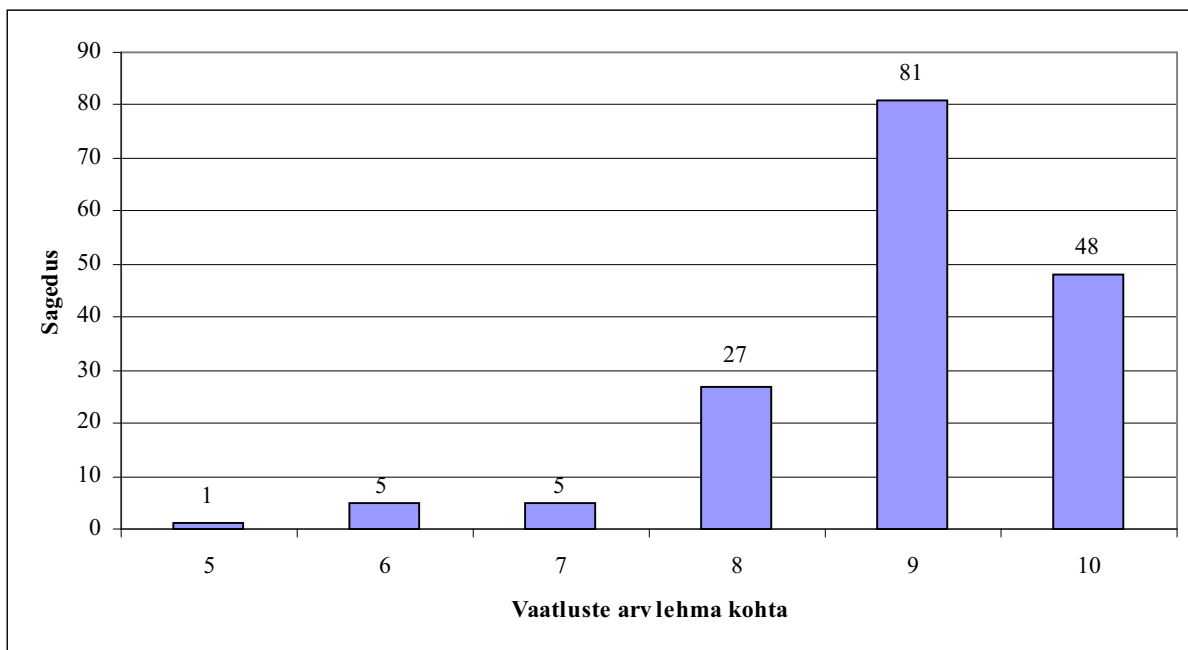
$$ECM_KG = piim * \frac{(383 * rasv + 242 * valk + 783.2)}{3140},$$

kus *piim* on kilogrammides päeva kohta ning *rasv* ja *valk* on protsendilise sisaldusena piimas.

Lehmaspetsiifiliste laktatsioonikõverate abil prognoosime ECM-toodangu 7. – 150. laktatsioonipäevale. Huvipakkuva perioodi piiramine 150. laktatsioonipäevaga baseerub varasemate aastate piimatoodangulastel uuringutel. Nende teadmiste põhjal on alust arvata, et lehma füsioloogiast tingitult on muutused energiabilansis suuremad just nimetatud perioodi vältel. Teadmised energiabilansi kohta laktatsiooni esimesel poolel võimaldavad lehmade söötmist optimaalsemaks muuta.

4.2.1. Andmestiku kirjeldus

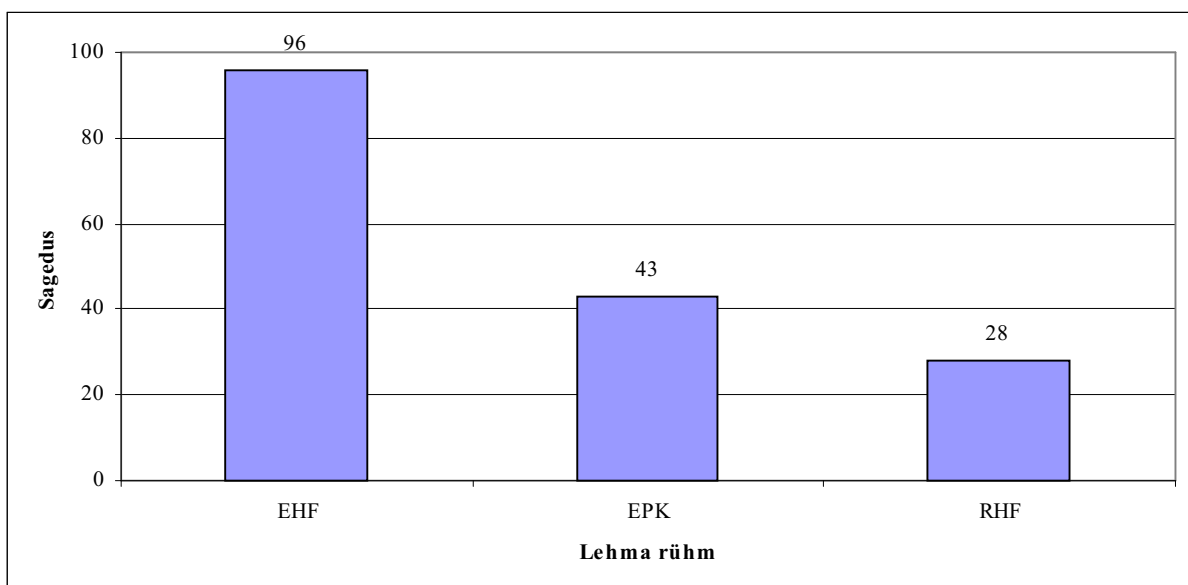
Andmed on kogutud ajavahemikul 23.10.2000 – 23.01.2005 kaks korda kuus Põlula katselaudas. Andmestik sisaldab 167 lehma andmeid esimese laktatsiooni 7. – 150. laktatsioonipäeva kohta. Esimestel päeval peale poegimist on lehma piima koostis tavapärasest erinev ning seetõttu pole esimesel nädalal peale poegimist sooritatud kontrolllõpside tulemused analüüsi kaasatud. Vaatlusi on kokku 1495. Igal lehmäl on vaadeldaval perioodil sooritatud 5 – 10 mõõtmist. Jooniselt 4.1 on näha, kui mitmel lehmäl konkreetne mõõtmiste arv esines.



Joonis 4.1. Erinevate vaatlussagedustega lehmade arv

Peaaegu pooltel lehmadel on uuritava perioodi jooksul üheksa mõõtmist. Teistest rohkem on ka kümne vaatlusega lehmi.

Andmestikus olevad lehmad on jagatud kolme rühma: eesti holstein (EHF), eesti punane kari (EPK), punasekirju holstein (RHF). Lehmade rühmadesse jagunemine on graafiliselt esitatud joonisel 4.2.



Joonis 4.2. Lehmade arv erinevates tõurühmades

Analüüsitava andmestik sisaldab järgmisi tunnuseid:

REG_NR – lehma registreerimisnumber;

BREED – lehma rühm;

FATHER – lehma isa nimi;

DIM – laktatsioonipäev;

ECM_KG – energia järgi korrigeeritud piim (ECM) kilogrammides.

Uuritavat tunnust kirjeldavad arvkarakteristikud on toodud alljärgnevas tabelis 4.1.

Tabel 4.1. ECM kirjeldavad arvkarakteristikud

Keskm	Standardh	Mediaan	Min	Max
29.092	5.051	29.030	7.813	47.011

4.2.2. Mudelid ja kovariatsioonistruktuuri sobitamine

Mudelite analüüsiks kasutame SAS protseduuri MIXED ja lähtume korduvate mõõtmiste analüüsi skeemist nimetatud protseduuriga [28]. Selle kohaselt on esialgse mudeli kirjapaneku järel vaja leida andmetega kooskõlas olev kovariatsioonistruktuur. Seejuures peaks arvestama heterogeensete dispersioonidega erinevate tõugude vahel ning ka sellega, et lehma ajas lähestikku asuvad vaatlused on omavahel rohkem korreleeritud kui üksteisest pikema ajaperioodi kaugusel olevad vaatlused. Seetõttu võiks oletada, et sobivamad on Toeplitzi **TOEP** ja esimest järku autoregressiivne **AR(1)** kovariatsioonistruktuurid kujul

$$\mathbf{TOEP} = \begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \cdots & \sigma_n \\ \sigma_1 & \sigma^2 & \sigma_1 & \cdots & \sigma_{n-1} \\ \sigma_2 & \sigma_1 & \sigma^2 & \cdots & \sigma_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_n & \sigma_{n-1} & \sigma_{n-2} & \cdots & \sigma^2 \end{pmatrix}, \quad \mathbf{AR(1)} = \begin{pmatrix} \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 & \cdots & \sigma^2 \rho^n \\ \sigma^2 \rho & \sigma^2 & \sigma^2 \rho & \cdots & \sigma^2 \rho^{n-1} \\ \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 & \cdots & \sigma^2 \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 \rho^n & \sigma^2 \rho^{n-1} & \sigma^2 \rho^{n-2} & \cdots & \sigma^2 \end{pmatrix},$$

kus σ^2 on mõõtmistele vastav dispersioon, ρ on autoregressiivsuse parameeter vaatluste omavahelise seose tugevuse määramiseks. Toeplitzi kovariatsioonistruktuur on üldisem,

kuna erinevate ajavahemike jaoks leitakse nendevaheline seos sõltumatult, mitte ühise parameetri abil.

Lisaks vaatame ka kovariatsioonistruktuure **CS** (näites 1.2 toodud V_i struktuur) ja **VC** kujul

$$\mathbf{CS} = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \cdots & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \cdots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \cdots & \sigma_1^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \cdots & \sigma^2 + \sigma_1^2 \end{pmatrix} \text{ ja } \mathbf{VC} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

Kovariatsioonimaatriksi struktuurid **CS** ja **VC** on omakorda lihtsustatud erijuhud eelnevalt kirjeldatud struktuuridest, eeldades konstantset vaatlustevahelist kovariatsioonistruktuuri **CS** puhul ning vaatluste vahelist sõltumatust struktuuri **VC** korral.

Vaadeldavate kovariatsioonistruktuuride võrdluseks antud mudeli puhul kasutame mudeli headust hindavaid statistikuid AIC (Akaike's Information Criterion), BIC (Schwarz's Bayesian Criterion) ja -2RLL (REML Log Likelihood). Nende põhjal tuleks otsustada struktuuri kasuks, mille puhul mudeli headuse statistikute väärtused on minimaalsed. Kahe erineva kovariatsioonistruktuuriga mudeli olulist erinevust kontrollime statistiku $[(-2LL_1) - (-2LL_2)]$ abil, kõrvutades seda χ^2 jaotusega. Vabadusastmete arv on mudelite kovariatsiooniparameetrite arvu vahe. Kui erinevus pole statistiliselt oluline, tuleb eelistada lihtsama kovariatsioonistruktuuriga mudelit.

Proovime piimatoodangu ECM laktatsioonikõvera kuju sobitamisel erinevaid juhuslike kordajatega funktsioone, mida on kasutatud varasemates uuringutes mujal maailmas.

Kuna kõik lehmad on pärit samast katselaudast, siis tavapäraselt modelis arvestatavat karja efekti antud juhul pole. Küll aga tuleks arvestada lehmade geneetiliste rühmade vahelist erinevust piimatoodangus. Seetõttu lisame mudelisse fikseeritud efektina rühma tunnuse *BREED*.

Mudel 1

Sobivaima mudeli otsinguid alustame lihtsaimast mudeli kujust, mis võiks piimatoodangu muutumist kirjeldada – laktatsioonipäeva kuupmudel:

$$ECM_KG = \mu + a_0 * BREED + a_1 * DIM + a_2 * DIM^2 + a_3 * DIM^3 + \varepsilon.$$

Selgub, et antud mudelit pole võimalik hinnata mitte ühegi valitud kovariatsioonistruktuuriga, sest arvutusprotsess ei koonu. Seetõttu asume otsima sobivamat mudelit.

Mudel 2

Kasutame ECM-toodangu hindamiseks punktis 3.2 toodud funktsiooni (3.8) ning lisame tõu fikseeritud efekti. Saame mudeli kujul

$$ECM_KG = \mu + a_0 * BREED + a_1 * C + a_2 * C^2 + a_3 * \ln(1/C) + a_4 * [\ln(1/C)]^2 + \varepsilon,$$

kus $C = DIM / 150$.

Tabel 4.2. Kovariatsioonistruktuuride sobivus modelile 2

Kov struktuur	-2RLL	AIC	BIC
VC	.	.	.
CS	7971.5	7985.5	8007.3
AR(1)	7964.5	7978.5	8000.4
TOEP	.	.	.

Tabelis 4.2 on toodud statistikud vaid kahele kovariatsioonistruktuurile, kuna **VC** ja **TOEP** struktuuri ei õnnestunud vaadeldava mudeliga analüüsitavatele andmetele sobitada.

Parimaks osutub esimest järku autoregressiivne kovariatsioonistruktuur.

Mudel 3

Uurime artiklis [8] esitatud laktatsioonikõverat modelleerivat kolmanda astme normeeritud Legendre polünoomi

$$ECM_KG = \mu + a_0 * BREED + a_1 * P'_0(t) + a_2 * P'_1(t) + a_3 * P'_2(t) + a_4 * P'_3(t) + \varepsilon,$$

kus $P'_i(t)$ on vastavalt järku i Legendre polünoom ja aeg t tähistab standardiseeritud laktatsioonipäeva:

$$t = \frac{2 * (DIM - DIM_{\min})}{DIM_{\max} - DIM_{\min}} - 1 = \frac{2 * DIM - 14}{150 - 7} - 1 = \frac{2 * DIM - 157}{143}.$$

Lähemalt Legendre polünoomidest punktis 3.2. Ülevaate erinevate struktuuride sobivusest annab järgnev tabel 4.3.

Tabel 4.3. Kovariatsioonistruktuuride sobivus mudelile 3

Kov struktuur	-2RLL	AIC	BIC
VC	.	.	.
CS	7945.0	7959.0	7980.8
AR(1)	7937.6	7951.6	7973.4
TOEP	7929.7	7961.7	8011.6

Sarnaselt mudelile 2 ei õnnestu ka mudeli 3 puhul hinnata kovariatsioonistruktuuri **VC**. Mudeli headuse karakteristikute AIC ja BIC põhjal on parim esimest järku autoregressiivne kovariatsioonistruktuur. Samas on statistiku -2LL põhjal parem **TOEP** struktuur. Kuna **AR(1)** ja **TOEP** struktuurid pole statistiku $[(-2LL_{AR(1)}) - (-2LL_{TOEP})]$ väärtuse põhjal oluliselt erinevad (statistiku väärtus 7.9; vabadusastmeid $16-7=9$), siis eelistame **AR(1)** kovariatsioonistruktuuriga mudelit, kui lihtsamat. Kovariatsioonistruktuuridel **CS** ja **AR(1)** on parameetreid ühepalju (7), seega paremaks võib ka selles võrdluses pidada väiksema statistiku -2LL väärtusega **AR(1)** struktuuri.

Mudel 4

Kasutame antud mudelis funktsiooni (3.9), mis on kompromiss toodud mudeli 3 ja Wilminki kõvera vahel:

$$ECM_KG = \mu + a_0 * BREED + a_1 * P_0(t) + a_2 * P_1(t) + a_3 * P_2(t) + a_4 * \exp(-0.05 * t) + \varepsilon,$$

kus $P_i(t)$ on i -ndat järku Legendre polünoom.

Sarnaselt mudelile 2, pole antud mudeli puhul võimalik sobitada andmetele kovariatsioonistruktuure **VC** ja **TOEP** (tabel 4.4). Mudeli headuse statistikute põhjal eelistame esimest järku autoregressiivset kovariatsioonistruktuuri.

Tabel 4.4. Kovariatsioonistruktuuride sobivus modelile 4

Kov struktuur	2RLL	AIC	BIC
VC	.	.	.
CS	7769.0	7783.0	7804.8
AR(1)	7763.6	7767.6	7789.4
TOEP	.	.	.

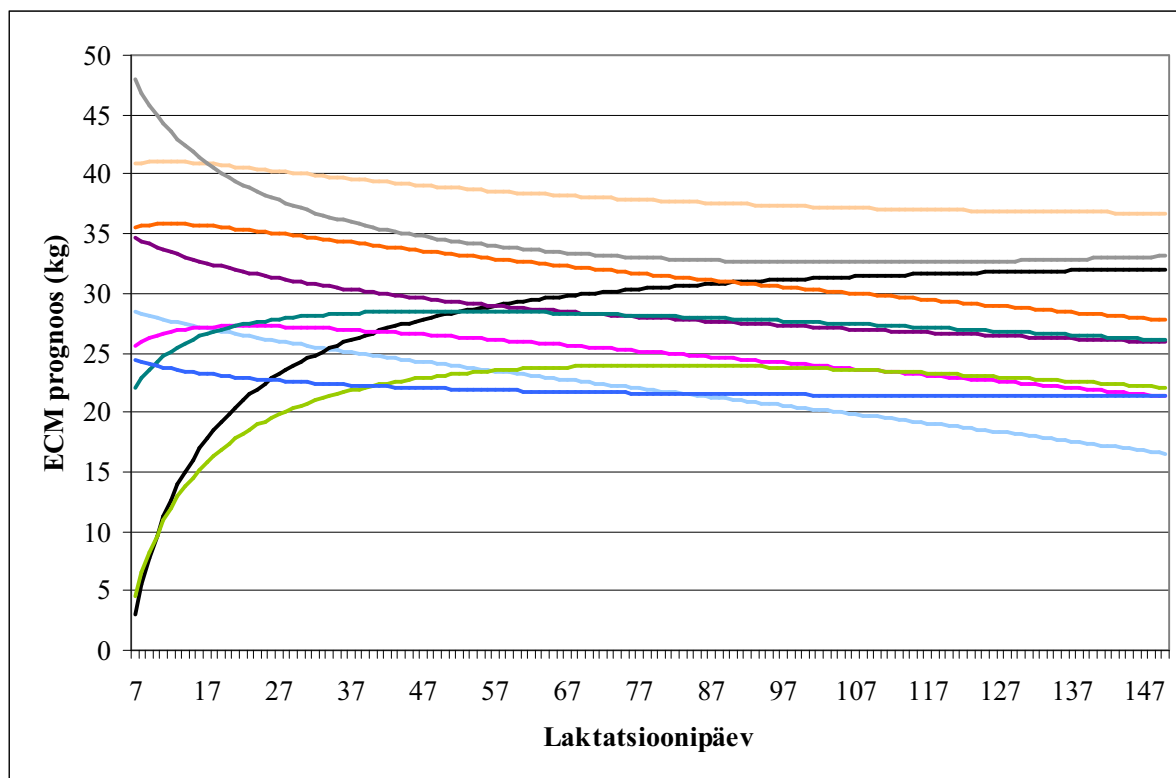
Seega osutus kõigi modelite puhul parimaks autoregressiivne kovariatsioonistruktuur **AR(1)**, mis sobib ka interpretatsiooni poolest.

4.2.3. Prognoosid ja parima mudeli valik

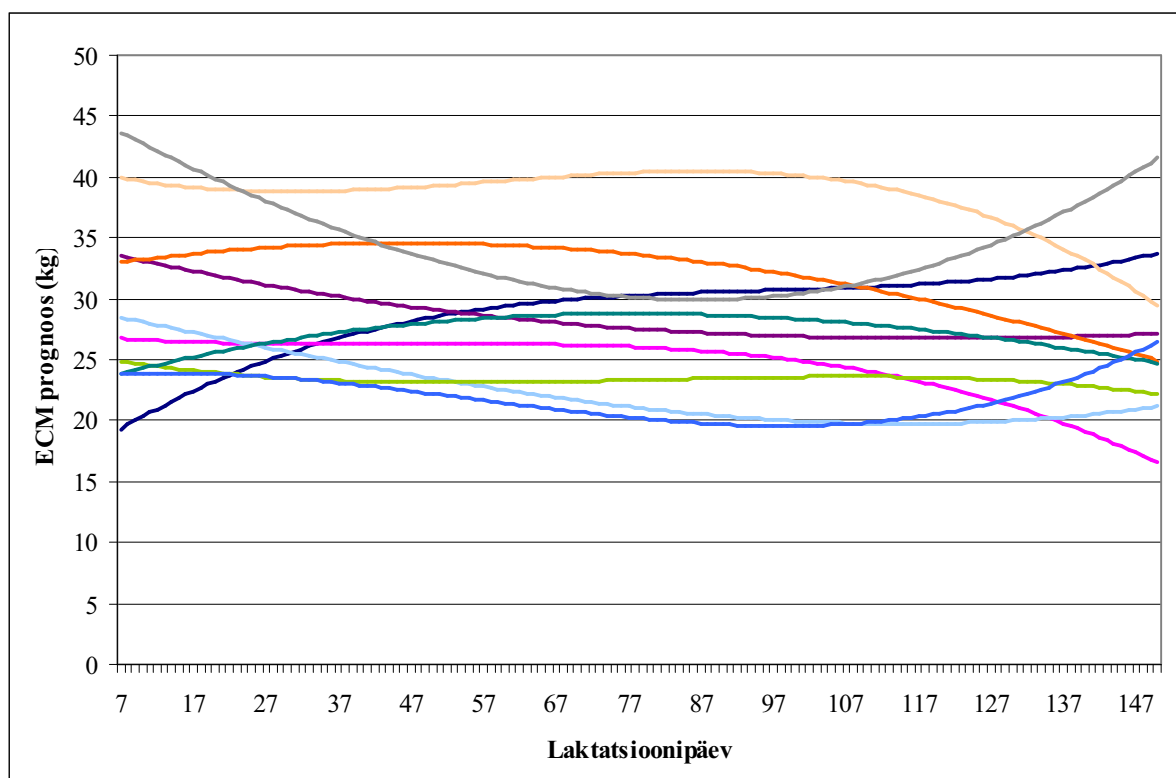
Olles määranud vaadeldavatele modelitele kovariatsioonistruktuuri, tuleb teha lõplik valik modelite 2, 3 ja 4 vahel. Parim mudel peaks võimalikult täpselt prognoosima individuaalseid laktatsioonikõveraid, st olema kooskõlas reaalsete andmetega. Samas ei tohi ta minna vastuollu üldtunnustatud ja füsioloogiliselt põhjendatud jõudlusnäitajate muutumistele ajas. Et seda hinnata, prognoosime võrreldavate modelitega laktatsioonikõverad.

Prognoosimiseks kasutame SAS protseduuri MIXED poolt väljastatud andmestikke fikseeritud ning juhuslikele efektidele (SAS väljundid on toodud lisas 2). Nii saab iga konkreetse lehma jaoks leida fikseeritud ja individuaalse efekti summeerimisel ning vastava laktatsioonipäeva väärtuse abil igapäevased ECM-toodangud.

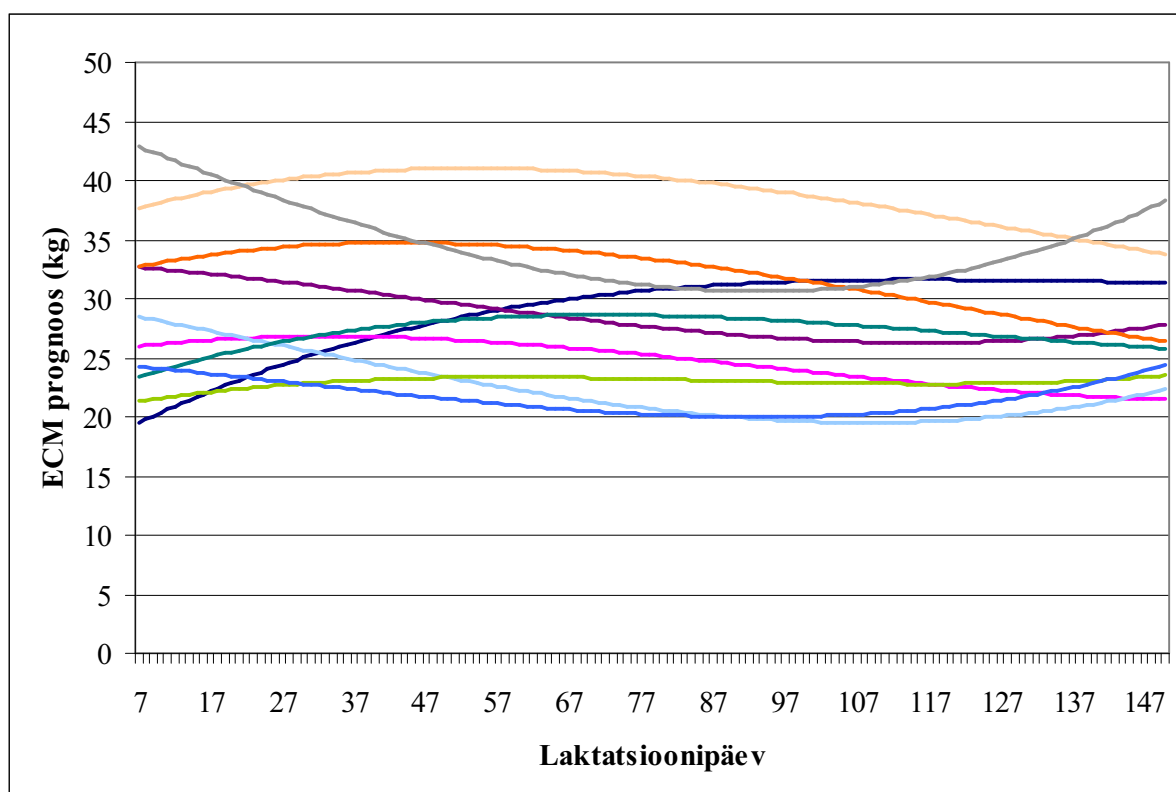
Igapäevased lehmaspetsiifilised ECM prognoosid leiame SAS protseduuri IML abil, fikseeritud ja juhuslike efektide ning laktatsioonipäeva ja rühma määravaid maatrikseid kasutades. Illustreerimaks saadut, on järgneval joonistel kujutatud valitud lehmade laktatsioonikõveraid, mis on hinnatud kolme erineva mudeli abil.



Joonis 4.3. Mudeli 2 abil prognoositud ECM laktatsioonikõverad erinevatele lehmadele



Joonis 4.4. Mudeli 3 ECM laktatsioonikõvera prognoosid erinevatele lehmadele



Joonis 4.5. Mudeli 4 abil leitud ECM prognoosikõverad erinevatele lehmadele

Kolmanda ja neljanda mudeli prognoosid on üsna sarnased, kuna mudelites kasutatavad funktsioonid sisaldavad suuremas osas sarnaseid komponente. Mainitud mudelite erinevus võrreldes mudeliga 2 on suurim just laktatsiooni alguspäevadel, kus mudeli 2 abil saadud prognoosid on teistega võrreldes suurema hajuvusega. Hiljem on mudel 2 aga stabiilne. Jooniste järgi otsustades tunduvad paremad prognoosid olema mudelitel 3 ja 4. Nende kahe mudeli vahel pole aga visuaalselt võimalik ühe või teise kasuks otsustada.

Uurime kõigepealt mudelite kirjeldavust juhuslikule veale vastava dispersioonikomponendi abil – mida väiksem on komponendi väärtus, seda suurem osa energia järgi korrigeeritud piimatoodangu varieeruvusest õnnestub meil vastava mudeliga kirjeldada. Kõige suurem on juhusliku vea dispersioonikomponent mudelil 2 (6.1276), mudelid 3 ja 4 on selles osas üsna sarnased (vastavalt 5.2804 ja 5.6891). Seega võib teistest paremaks pidada mudelit 3.

Edasi võrdleme prognooside kooskõla tegelikult mõõdetud väärtustega. Selleks leiame Pearsoni korrelatsioonikordaja ja keskmise ruutvea (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2,$$

kus $\hat{y}_i$ on prognoositud ja y_i tegelik ECM väärtus laktatsioonipäeval i ning n tähistab vaatluste arvu.

Tulemused on toodud järgnevas tabelis 4.5.

Tabel 4.5. Erinevate mudelite sobivusnäitajad

Mudel	Korrelatsioon	MSE
M2	0.9056	4.6967
M3	0.9348	3.4035
M4	0.9158	4.2242

Nagu näha, on kõikide mudelite puhul vaadeldavad näitajad üsna ühesugused. Teistest veidi suurema korrelatsioonikordaja ja veidi väiksema keskmise ruutveaga on mudel 3.

Uurime ka seda, kui hästi on vaadeldavad mudelid prognoosinud ECM väärtusi igale lehmale eraldi. Selleks leiame korrelatsioonid lehma vaatluste ja vastavatele lüpsipäevadele arvutatud prognooside vahel ning vaatame saadud seoste tugevust (st korrelatsioonikordaja absoluutväärtust). Sellest annab ülevaate tabel 4.6.

Tabel 4.6. Lehmadele leitud korrelatsioonide tugevust kirjeldavad näitajad

Mudel	Mediaan	Min	Max	Al kvartiil	Ül kvartiil
M2	0.6431	0.0043	0.9748	0.4045	0.8341
M3	0.7809	0.0748	0.9891	0.6411	0.8867
M4	0.7073	0.0432	0.9852	0.4964	0.8461

Taas on sobivuse poolest teistest veidi parem mudel 3 – minimaalne ja maksimaalne korrelatsioon lehma tegelike ja prognoositud ECM väärtuste vahel on teistest suuremad. Kõige suurem erinevus vaadeldavate mudelite hulgas paistab silma alumise kvartiili väärtuses, mis on mudeli 3 puhul teistest märgatavalt suurem ning viitab sellele, et tugevamaid korrelatsioone esineb selle mudeli puhul teistest rohkem.

Artiklis [8] võrreldakse erinevaid mudeleid piimatoodangu laktatsioonikõvera kirjeldamiseks. Mudelite 3 ja 4 sobitamisel juhuslikule osale isa- ja loomamudelid leidsid artikli autorid, et parem on mudel 3. Sobivuskriteeriumina kasutati -2RLL minimaalset väärtust. Üksnes sellest kriteeriumist lähtudes on meil eelistatum mudel 4.

Eelnev võrdlus põhineb mudelite matemaatilisel headusel. Tulemuste põhjal võiks parimaks pidada mudelit 3, millest veidi halvemad näitajad on mudelil 4. Lisaks konkreetsete reaalsete väärtustega kooskõla saavutamisele on aga tähtis ka mudeli interpreteeritavus – kui hästi vastavad prognoositud kõverad praktikas kinnitust leidnud laktatsioonikõverale. Oodatava piimatoodangu laktatsioonikõvera kohaselt peaks perioodi algul toodang kasvama ning saavutama maksimumi teiseks laktatsioonikuuks, hakates seejärel vähehaaval langema [29]. Tuleb silmas pidada aga seda, et Põlula katselauda lehmad ei esinda Eesti keskmist lehmade populatsiooni. Paljud sealsed lehmad on kõrge toodanguga ja neil esineb suuri väärtusi kogu vaadeldava perioodi vältel. Muutlikke vaatlusi lehma kohta esineb samuti üsna palju. Viimaseid polegi võimalik tuvastada erindite otsimisel andmestikust, sest üksikvaatlusena on nad normaalsed. Küll aga ilmneb erandlik trend konkreetse looma järjestikuseid mõõtmisi kõrvutades. Nagu näha joonistelt 4.4 ja 4.5, on juhuslike regressioonikordajatega mudelid küllalt paindlikud modelleerimaks ka erandliku laktatsiooni kulgemisega lehmade toodanguid ning tulemuseks võivad olla oodatule vastupidise kujuga kõverad. Seega võib siinkohal andmestiku uuesti üle vaadata, erandlikud lehmad välja jätta ning sobitada mudelit uuesti. Põlula katse puhul on tegu niigi väikese valimiga ning seetõttu püüame pigem leida mudelit, mis ka erandlike lehmade toodangutele enam loogilised trajektoolid prognoosiks. Toodud kaalutlustest lähtuvalt võiks antud juhul paremaks tunnistada mudeli 4, mille puhul pole prognoositud laktatsioonikõveratel uuritava perioodi lõpus nii järsku tõusu.

Võttes kokku statistilise sobivuse ja kooskõla reaalsusega võib ECM-toodangu prognoosimisel teistest paremaks pidada mudelit 4.

Leitud prognoosidest saame igale lehmale välja lugeda maksimaalse ECM-toodangu ja sellele vastava laktatsioonipäeva. Tulevikus on kavas leida lisaks ECM-toodangu prognoosidele individuaalsed kõverad ka lehmade kehakaalule ja söömusele. Nimetatud kolme näitaja alusel saab seejärel arvutada igapäevaselt energiabilansi trajektoolid igale lehmale eraldi.

4.2.4. Pullide ECM-toodangu aretusväärtuste hindamine leitud prognooside abil

Saadud parima mudeli prognoose on meil võimalik kasutada ka geneetiliseks hindamiseks, et välja selgitada, kui hea ECM toodangu pärandavad pullid andmestikus olevatele lehmadele kui oma järglastele.

Laktatsioonikõverate abil leitud ECM prognoose saame kasutada isade ECM toodangu aretusväärtuse arvutamiseks igale laktatsioonipäevale i eraldi. Selleks leiame iga pulli järglaste keskmise toodangu erinevuse populatsiooni üldkeskmisest:

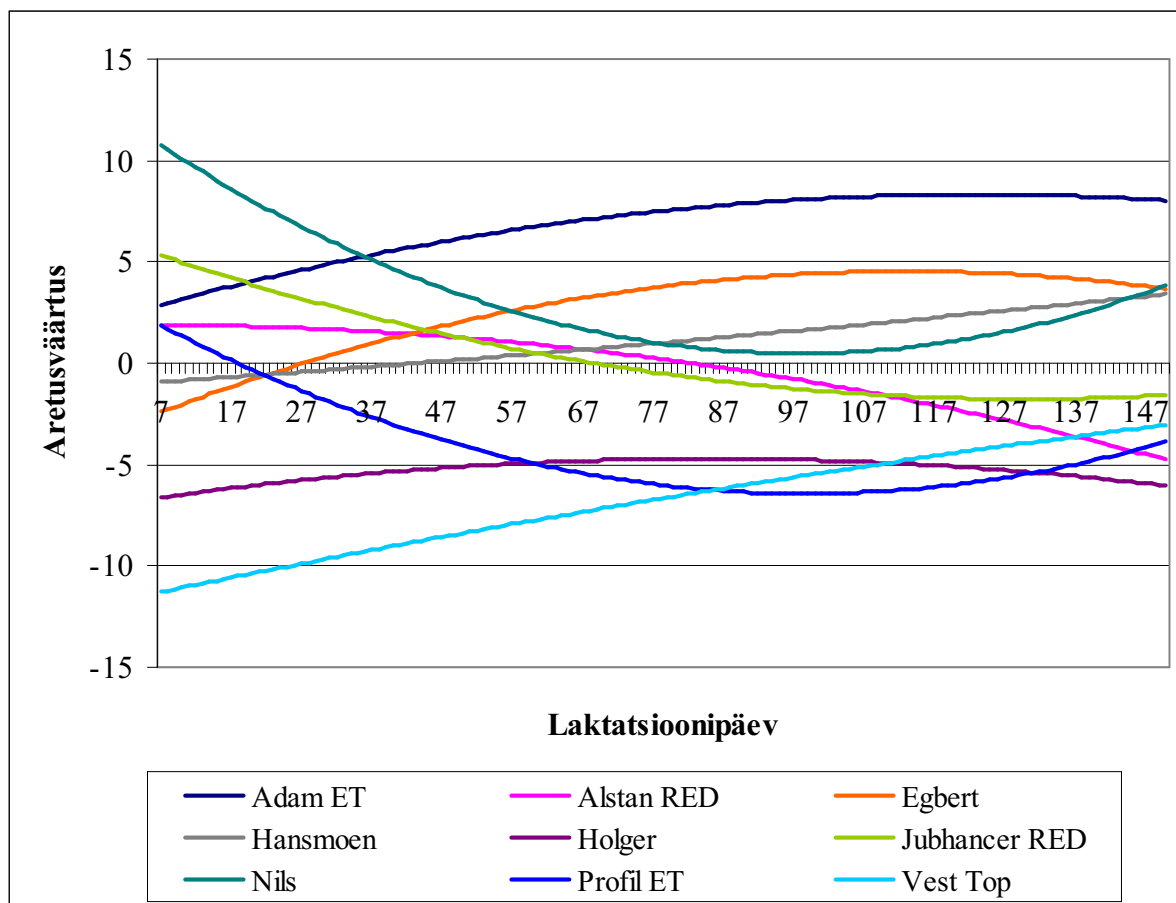
$$\hat{A}_{s_i} = 2(\bar{P}_{s_i} - \bar{P}) \quad (4.1)$$

kus $\hat{A}_{s_i}$ on pulli s aretusväärtus laktatsioonipäeval i , $\bar{P}_{s_i}$ on isa s järglaste keskmine uuritava tunnuse väärtus laktatsioonipäeval i ja $\bar{P}$ populatsiooni üldkeskmine [30]. Aretusväärtus näitab seega, kui palju üle või alla populatsiooni keskmise taseme jääb uuritava tunnuse väärtus konkreetsel indiviidil ning oleneb seega indiviidist ja populatsioonist.

Tabel 4.6. Erineva arvu tütardega pullide arv

Pulli tütarde arv	1	2	3	4	5	6	7	8	9	10
Pullide arv	37	12	8	4	2	3	1	1	1	1

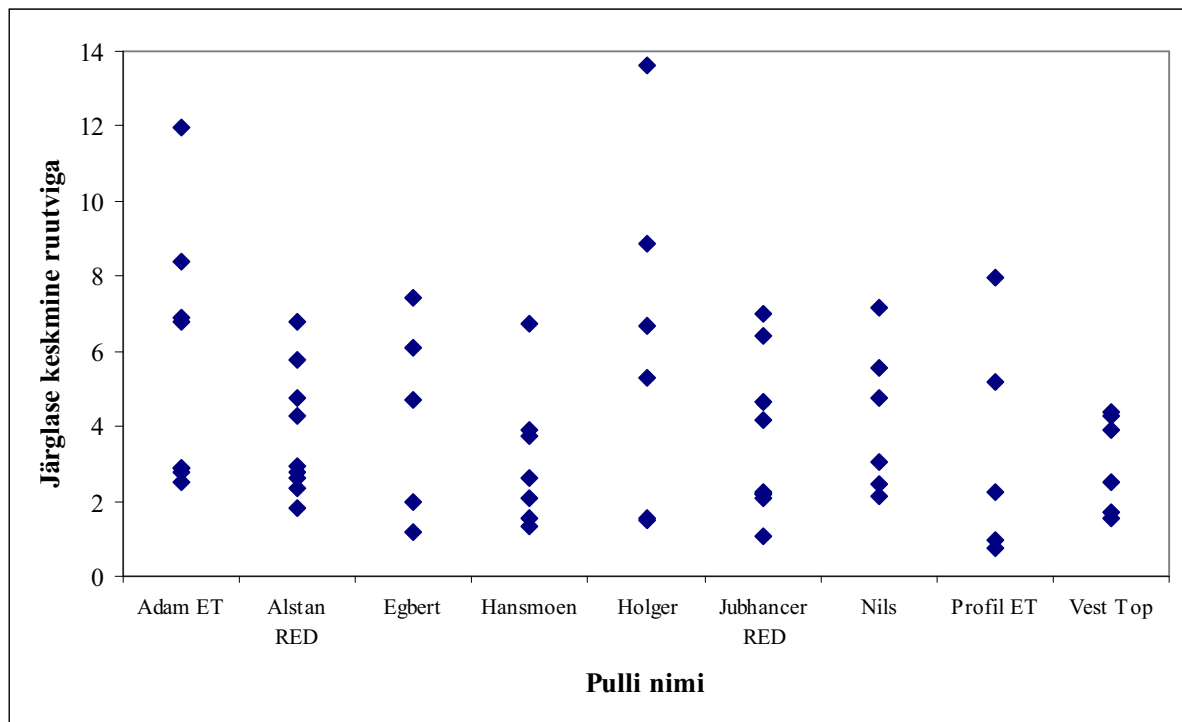
Nagu näha tabelist 4.6, on enamikul pullidest vaadeldavas andmestikus vaid üks järglane. Neile pullidele pole otstarbekas aretusväärtusi leida. Mida rohkem järglasi aretusväärtuse arvutamiseks kasutada on, seda usaldusväärsemad on tulemused. Joonisel 4.6 on kujutatud vähemalt viit järglast omanud pullide valemi (4.1) alusel hinnatud aretusväärtused.



Joonis 4.6. Erinevate pullide hinnatud aretusväärtused laktatsiooni esimesel poolel

Üsna sarnase kujuga kõverad on pullidel Nils ja Profil ET. Nende tütarde toodang näitab laktatsiooni algul langustrendi võrreldes andmestiku keskmisega. Vastupidist trendi näitavad omavahel sarnased kõverad pullidele Egbert ja Adam ET. Teistest veidi suurema aretusväärtuse poolest paistab silma Adam ET ning üsna stabiilselt madala aretusväärtusega on Holger. Vest Topi järglaste laktatsiooni alguse madal toodang näitab tõusutrendi.

Järgnevalt uurime, kui täpne on sobivaimaks tunnistatud juhuslike regressioonikordajatega mudel 4 antud pullide puhul. Kas mõne pulli järglastele sobib see mudel paremini? Selleks leiame iga pulli järglaste keskmise ruutvea, mis on toodud joonisel 4.7.



Joonis 4.7. Erinevate pullide tütarde ECM toodangu ruutkeskmised vead

Enamikel pullidel on tütarde keskmiste ruutvigade hajuvus sarnane. Teistest suuremaid väärtusi esineb pullidel Holger ja Adam ET. Väiksemad keskmised ruutvead on pullil Vest Top. Siiski pole erinevused pullide vahel kuigi suured ning andmete vähesuse tõttu pole alust väita, et mudel 4 sobib teatud pullide järglaste ECM-toodangu prognoosimiseks paremini ja teatud pullide järglaste ECM-toodangu prognoosimiseks halvemini.

Leiame antud pullidele ka vaadeldava perioodi aretusväärtuse, mis on igapäevaselt leitud aretusväärtuste summa:

$$\hat{A}_s = \sum_{i=7}^{150} \hat{A}_{s_i} .$$

Saadud summaarsete aretusväärtuste alusel saame pullid reastada paremuse järgi (tabel 4.7).

Tabel 4.7. Pullide paremusjärjestus aretusväärtuse järgi

Pull	Aretusväärtus
Adam ET	983,1
Nils	446,9
Egbert	393,2
Hansmoen	157,7
Jubhancer RED	41,6
Alstan RED	-51,7
Profil ET	-622,2
Holger	-758,3
Vest Top	-983,1

Seega võime öelda, et laktatsiooni esimese poole energia järgi korrigeeritud piima toodang on kõige suurem pulli Adam ET järglastel. Paremuselt teine on pull Nils. Jooniselt 4.6 on näha, et nimetatud pullidel on aretusväärtus positiivne kogu vaadeldava perioodi jooksul. Suuruselt kolmanda aretusväärtusega pullil Egbert on aretusväärtus esimesel laktatsioonikuul negatiivne.

Kokkuvõte

Käesolevas magistritöös anname ülevaate juhuslike kordajatega üldise lineaarse segamudeli olemusest ning rakendustest loomade aretuses, kus nimetatud mudelit kasutatakse järjest enam nii majanduslikult olulisi tunnuseid mõjutavate faktorite välja selgitamisel kui ka loomade geneetilisel hindamisel. Eelkõige viimase eesmärgi tarvis väljatöötatud kontrollpäeva mudeli fikseeritud regressioonikordajatega varianti rakendatakse ka Eestis Jõudluskontrolli Keskuses veiste jõudlustunnuste aretusväärtuste hindamiseks. Lisaks kontrollpäeva mudelile leiab juhuslike kordajatega üldine lineaarne segamudel kasutust loomade aretuses ka näiteks kasvukõverate, elukestuse ning genotüübi ja keskkonna koosmõju alastes uuringutes. Enamikes neist valdkondadest on antud tüüpi mudelite rakendamine alles algusjärgus ning sobiva mudeli väljakujunemine nõuab veel täiendavaid uuringuid.

Töös rakendame juhuslike kordajatega üldist lineaarset segamudelit rakendusuuringu “Eesti veisetõugude maksimaalse piimajõudluse väljaselgitamine” andmetele. Selle uuringu mitmekesiste näitajate hulgast pakub antud juhul huvi 7. – 150. laktatsioonipäeva energia järgi korrigeeritud piimatoodang. Eesmärk on leida andmeid hästi kirjeldav ning ka reaalsusega kooskõlas olev mudel lehmadele individuaalsete laktatsioonikõverate hindamiseks. Selleks sobitame antud andmetele varasemates aretusalaates uuringutes kirjeldatud laktatsioonikõvera funktsioone. Andmete analüüsimiseks kasutame SAS protseduuri MIXED.

Parima mudeli väljaselgitamist alustame mudelitele sobiva kovariatsioonistruktuuri määramisest ning seejärel võrdleme mudelite headust jääkvarieeruvuse, keskmise ruutvea ning prognooside ja tegelike väärtuste vaheliste korrelatsioonide abil. Kuna mudelite headuse näitajates väga suuri erinevusi ei esine, teeme lõpliku otsuse oodatava kõvera kujuga paremini sobiva mudeli kasuks. Vastav mudel on kujul

$$ECM_KG = \mu + a_0 * BREED + a_1 * P_0(t) + a_2 * P_1(t) + a_3 * P_2(t) + a_4 * \exp(-0.05 * t) + \varepsilon,$$

kus $P_i(t)$ on i -järku Legendre polünoom, $BREED$ on lehma rühma fikseeritud mõju, ε on juhuslik viga ning t on lõigule $[-1;1]$ standardiseeritud ajahetk:

$$t = \frac{2 * (LP - LP_{\min})}{(LP_{\max} - LP_{\min})} - 1,$$

kus $LP_{\min}$ on varaseim ajahetk ning $LP_{\max}$ on hiliseim andmetes esinev ajahetk.

Toodud mudeli põhjal leitud individuaalsete laktatsioonikõverate abil leiame andmestikus olevate lehmade isadele (kellel andmestikus on vähemalt viis järglast) aretusväärtused nii igapäevaselt kui ka vaadeldava perioodi summaarse väärtusena. Vastavatelt kõveratelt on näha, et enamike pullide puhul muutub aretusväärtus erinevalt. Siiski ei saa järglaste keskmiste ruutvigade põhjal väita, et valitud mudel prognoosiks mõne pulli tütarde energia järgi korrigeeritud piimatoodangut paremini. Huvi pakkunud perioodi summaarse aretusväärtuse põhjal saame vaatlusalused pullid järjestada geneetilise paremuse alusel. Suurima aretusväärtusega on pull Adam ET, talle järgnevad pullid Nils ja Egbert.

Kokkuvõttes saime juhuslike kordajatega üldise lineaarse segamudeli abil individuaalsed laktatsioonikõverad lehmade ECM-toodangule ning hindasime saadud tulemuste alusel ka pulle nende poolt oma tütardele pärandatava ECM-toodangu geneetilise potentsiaali alusel. Tulevikus on kavas leida lisaks ECM-toodangu prognoosidele individuaalsed kõverad ka lehmade kehakaalule ja söömusele, misjärel saab nimetatud kolme näitaja alusel läbi viia energiabilansi alaseid uuringuid.

Random regression model in animal breeding

Master Thesis

Mirjam Vallas

Summary

The present paper provides an overview of the applications of general linear mixed model in animal breeding. In the present context general linear mixed models are also called random regression models. The methodological overview is illustrated with an example on modelling the data of a study on “The maximal milk yield of Estonian cattle breed”. Among various variables measured in the study energy-corrected milk yield is of special interest.

The aim of the data modelling part is to find a good reality-reflecting fit for assessing individual lactation curves of cattle within 7 to 150 days of lactation. For this purpose functions for lactation curves proposed in the literature were used. Data analysis was implemented in software package SAS/STAT (version 8.2) using procedure MIXED. The final fit, not statistically the best, but the best considering the accordance with reality, is the following:

$$ECM_KG = \mu + a_0 * BREED + a_1 * P_0(t) + a_2 * P_1(t) + a_3 * P_2(t) + a_4 * \exp(-0.05 * t) + \varepsilon ,$$

where $P_i(t)$ is the i -th order Legendre's polynomial, $BREED$ the fixed effect of cattle group, ε random error and t standardized time in the following form:

$$t = \frac{2 * (LP - LP_{\min})}{(LP_{\max} - LP_{\min})} - 1 ,$$

where $LP_{\min}$ is the earliest time point and $LP_{\max}$ the latest time point present in the data.

Based on the individual lactation curves obtained from the fitted model the breeding value, both daily and considering it as a total measure over the observed period, for sires having at least five offspring were estimated. The total breeding value enables to order the considered sires from the best to the worst.

To conclude, applying general linear mixed models the individual lactation curves for each cow's ECM-milk yield were obtained. The results were used to assess the sires comparing the genetic potential for good ECM-milk yield inherited by their offspring.

In addition to the ECM-milk yield considered here, the individual curves for feed intake and cow's weight will be assessed as an extension of the present work. Then, based on the three aforementioned measures the studies about energy balance may be performed.

Kirjandus

1. Searle, S.R., Casella, G., McCulloch, C.E., Variance components. Wiley, 1992.
2. Statistical models with random effects ,
[<http://statmaster.sdu.dk/st113/english/index.html>], 17.01.2005.
3. Hedeker, D. Longitudinal data analysis,
[<http://tigger.uic.edu/~hedeker/bstt513.html>], 19.01.2005.
4. Jõudluskontrolli Keskus,
[<http://www.jkkeskus.ee/page.php?page=1090>], 11.03.2005.
5. Saveli, O., Teinberg, R., Uba, M. Pullide aretusväärtuste hindamine järglaste jõudlusandmete alusel. Agraarteadus, 5(4), 1994, 456 – 462.
6. Eesti Jõudluskontrolli Aastaraamat 1996. Elmatar, 1997.
7. Schaeffer, L.R. Application of random regression models in animal breeding. Livestock Production Science, 86, 2004, 35 – 45.
8. Jakobsen, J.H. Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. Journal of Dairy Science, 85, 2002, 1607 – 1616.
9. Jensen, J. Genetic evaluation of dairy cattle using test-day models. Journal of Dairy Science, 84, 2001, 2803 – 2812.
10. van der Werf, J.H.J., Goddard, M.E., Meyer, K. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. Journal of Dairy Science, 81, 1998, 3300 – 3308.

11. Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari E.A., Meuwissen, T.H.E. Strategies for estimating the parameters needed for different test-day models. *Journal of Dairy Science*, 83, 2000, 1125 – 1134.
12. Swalve, H.H., Theoretical basis and computational methods for different test-day genetic evaluation methods. *Journal of Dairy Science*, 83, 2000, 1115 – 1124.
13. Szyda, J., Liu, Z. Modelling test day data from dairy cattle, [<http://gen.ar.wroc.pl/Tdmodel.pdf>], 10.04.2005.
14. Schenkel, F.S., Miller, S.P., Jamrozik, J., Wilton, J.W. Random regression analysis of weight gain of station-tested beef bulls. [<http://bru.aps.uoguelph.ca/Articles01/2001-pg5.pdf>], 5.04.2005.
15. Reents, R., Dopp, L. Genetic evaluation for dairy production traits with a test day model for multiple lactations. *Interbull Bulletin*, 14, 1996, 113 – 117.
16. Meyer, K., Hill, W.H. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated 'records by restricted maximum likelihood. *Livestock production science*, 47, 1997, 185 – 200.
17. Sherchand, L., McNew, R.W., Kellogg, D.W., Johnson, Z.B. Selection of a mathematical model to generate lactation curves using daily milk yields of holstein cows. *Journal of Dairy Science*, 78, 1995, 2507 – 2513.
18. Schaeffer, L.R., Dekkers, J.C.M. Random regressions in animal models for test-day production in dairy cattle. *Proceedings of the 5th world congress on genetics applied to livestock production*, 18, 1994, 443 – 446.
19. Jamrozik, J., Schaeffer, L.R. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation holsteins. *Journal of Dairy Science*, 80, 1997, 762 – 770.

20. Guo, Z., Swalve, H.H. Comparison of different lactation curve sub-models in test-day models. *Interbull Bulletin*, 16, 1997, 75 – 79.
21. Jamrozik, J., Schaeffer, L.R., Dekkers, J.C.M. Random regression models for production traits in Canadian Holsteins. *Interbull Bulletin*, 14, 1996, 124 – 134.
22. Jamrozik, J., Schaeffer, L.R., Liu, Z., Jansen, G. Multiple trait random regression test day model for production traits. *Interbull Bulletin*, 16, 1997, 43 – 47.
23. Kistemaker, G.J. The Canadian test day model using Legendre polynomials. *Interbull Bulletin*, 31, 2003, 202 – 204.
24. Coffey, M.P., Simm, G., Brotherstone, S. Energy balance profiles for the first three lactations of dairy cows estimated using random regression. *Journal of Dairy Science*, 85, 2002, 2669 – 2678.
25. Strabel, T., Ptak, E., Szyda, J., Jamrozik, J. Multiple-lactation random regression test-day model for Polish Black and White Cattle. *Interbull Bulletin*, 32, 2004, 133 – 136.
26. Lidauer, M., Mäntysaari, E.A. Multiple trait reduced rank random regression test-day model for production traits. *Interbull Bulletin*, 22, 1999, 74 – 80.
27. Strabel, T., Szyda, J., Ptak, E., Jamrozik, J. Comparison of random regression test-day models for production traits of dairy cattle in Poland. *Interbull Bulletin*, 31, 2003, 197 – 201.
28. Lehto, K. Retrospektiivsetele korduvmõõtmistele tuginevad laste ja noorukite kasvumudelid. *Magistritöö*. Tartu, 2003.
29. Roginski, H., Fuquay, J.W., Fox, P.F. *Encyclopedia of dairy sciences*, Vol 2. Elsevier Science, 2003.
30. Teinberg, R. *Põllumajandusloomade geneetika*. Valgus, 1978.

LISA1

Loomakasvatusalaste mõistete seletus

Aditiivgeneetiline efekt – indiviidi üjksikute geenide summaarsest mõjust tingitud erinevus populatsiooni keskmisest.

Aretusväärtus – kahekordne indiviidi järglaste keskmine hälve populatsiooni keskmisest vastava tunnuse osas.

Energiabilanss – söödast saadava ja piimatootmiseks ning enese elamiseks kuluva energiakoguse vahe.

Genotüüp – konkreetse organismi geenide kogum.

Kehaline konditsioon – ainete saamise ja äraandmise suhe looma ainevahetuses.

Laapumisomadused – juustu moodustumisprotsessi omadused.

Laktatsioon, laktatsiooniperiood – lüpsiperiood.

Metaboliit – ainevahetuse käigus tekkiv teatav ühend.

Päritavus - kvantitatiivse tunnuse populatsioonisisese muutlikkuse see osa, mis on tingitud genotüübilistest erinevustest indiviidide vahel.

Söödaväärindus – söödakulu toodanguühiku kohta.

Söömus – looma poolt tarbitud söötade, neis sisalduva kuivaine või orgaanilise aine kogus.

Uuslüpsiperiood – lüpsiperiood poegimisest tiinestumiseni.

LISA2

Peatükis 4 analüüsitud mudelite SAS väljatrükk

Mudel2

The Mixed Procedure

```
Breed2          3860205
                  3  EHF EPK RHF
```

Dimensions

Covariance Parameters	7
Columns in X	8
Columns in Z Per Subject	15
Subjects	167
Max Obs Per Subject	11
Observations Used	1495
Observations Not Used	64
Total Observations	1559

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	9021.34216181	
1	3	8129.40150146	3.95163760
2	2	8093.80751516	2.60570080
3	2	8070.78496534	4.08209456
4	2	8046.77372056	3.80761370
5	2	8035.13361685	1.10721731
6	2	8023.06060638	0.25424499
7	2	8014.42971483	0.07320361
8	2	7961.24665641	2.98965615
9	2	7944.24249693	3.07307762
10	2	7913.76304709	0.01468887
11	2	7896.54196022	0.01417645
12	2	7865.29120739	0.00893295
13	2	7845.93969513	0.00140086
14	1	7841.77305515	0.00016164
15	1	7841.31686989	0.00000558
16	1	7841.30225312	0.00000001

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	Reg_nr	Breed2 EHF	6.2475
AR(1)	Reg_nr	Breed2 EHF	0.1490
Variance	Reg_nr	Breed2 EPK	4.3846
AR(1)	Reg_nr	Breed2 EPK	0.1485
Variance	Reg_nr	Breed2 RHF	6.6788
AR(1)	Reg_nr	Breed2 RHF	0.1933
Residual			5.5736

Fit Statistics

-2 Res Log Likelihood	7841.3
AIC (smaller is better)	7855.3
AICC (smaller is better)	7855.4
BIC (smaller is better)	7877.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	1180.04	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Breed2	2	133	7.16	0.0011
P1	1	1350	41.13	<.0001
P2	1	1073	2.00	0.1575
EP	1	1174	29.58	<.0001

Mudel3

The Mixed Procedure

Breed2 3860205
 3 EHF EPK RHF

Dimensions

Covariance Parameters	7
Columns in X	8
Columns in Z Per Subject	15
Subjects	167
Max Obs Per Subject	11
Observations Used	1495
Observations Not Used	64
Total Observations	1559

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	9028.59198872	
1	3	8503.16386199	9.08436348
2	2	8433.87173588	17.98380057
3	2	8401.80803251	0.93535596
4	2	8363.76150623	3.27005058
5	2	8335.91695262	5.60577828
6	2	8314.14492986	2.22189346
7	2	8308.40963531	1.27154245
8	2	8301.86058525	0.45715045
9	2	8286.85905712	1.05102899
10	2	8268.83313871	1.28002455
11	2	8263.19884585	3.40080944
12	2	8194.31248399	11.63043867
13	2	8156.69937386	10.66599684
14	2	8135.97219300	0.04310452
15	2	8098.35378474	0.01852811
16	2	8059.17343238	0.02239838
17	2	7991.19028358	0.02685758
18	2	7967.11393249	0.01025738
19	2	7939.16451459	0.00054399
20	1	7937.62628396	0.00001325
21	1	7937.59103367	0.00000002
22	1	7937.59098142	0.00000000

Convergence criteria met.

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	Reg_nr	Breed2 EHF	5.3547
AR(1)	Reg_nr	Breed2 EHF	0.1461
Variance	Reg_nr	Breed2 EPK	3.4644
AR(1)	Reg_nr	Breed2 EPK	0.2446
Variance	Reg_nr	Breed2 RHF	5.2348
AR(1)	Reg_nr	Breed2 RHF	0.2597
Residual			5.2804

Fit Statistics

-2 Res Log Likelihood	7937.6
AIC (smaller is better)	7951.6
AICC (smaller is better)	7951.7
BIC (smaller is better)	7973.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	1091.00	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Breed2	2	172	8.97	0.0002
P1	1	706	13.62	0.0002
P2	1	714	6.93	0.0086
P3	1	721	4.64	0.0315

Model4

The Mixed Procedure

Breed2 3860205
 3 EHF EPK RHF

Dimensions

Covariance Parameters	7
Columns in X	7
Columns in Z Per Subject	15
Subjects	167
Max Obs Per Subject	11
Observations Used	1495
Observations Not Used	64
Total Observations	1559

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	9002.58465433	
1	3	7757.17954478	0.00118159
2	2	7753.81794496	0.00008994
3	1	7753.57913905	0.00000082
4	1	7753.57720711	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	Reg_nr	Breed2 EHF	7.8549
AR(1)	Reg_nr	Breed2 EHF	-0.4872
Variance	Reg_nr	Breed2 EPK	4.8834
AR(1)	Reg_nr	Breed2 EPK	-0.3395
Variance	Reg_nr	Breed2 RHF	7.5729
AR(1)	Reg_nr	Breed2 RHF	-0.1388
Residual			5.6891

Fit Statistics

-2 Res Log Likelihood	7753.6
AIC (smaller is better)	7767.6
AICC (smaller is better)	7767.7
BIC (smaller is better)	7789.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	1249.01	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Breed2	2	92.4	3.94	0.0229
t	1	1124	22.25	<.0001
P22	1	1124	21.77	<.0001
EP1	1	1124	22.24	<.0001