

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Markus Erik Tommula

**Kas ja kuivõrd ennustab piima rasvhappeline
koostis lehma tiinestumist?**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendaja: PhD Tanel Kaart

TARTU 2026

KAS JA KUIVÕRD ENNUSTAB PIIMA RASVHAPPELINE KOOSTIS LEHMA TIINESTUMIST?

Bakalaureusetöö

Markus Erik Tommula

Lühikokkuvõte

Bakalaureusetöö eesmärk on uurida, kas kontroll-lüpsil määratud piima rasvhappeline koostis võimaldab lehma tiinestumise prognoosimist esimesel seemendusel. Töö koosneb kolmest peatükist, esimene neist annab ülevaate rasvhapete määramisest ja päritolust, teine peatükk kirjeldab analüüsis kasutatud mudelit ja selle tõlgendust ning kolmas peatükk kirjeldab analüüsi tulemusi.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: Logistiline mudel, piimalehmad, rasvhapped, tiinestumine.

CAN MILK FATTY ACID COMPOSITION BE USED TO PREDICT COW FERTILITY?

Bachelor thesis

Markus Erik Tommula

Abstract

The aim of this bachelor's thesis is to analyse the predictive value of milk fatty acid composition, recorded during test milking, for cow fertilisation outcomes. The thesis consists of three chapters, the first of which provides a brief overview of the origin of fatty acids. The second chapter introduces the model used for analysis and its interpretation, and the third chapter presents the results of the data analysis.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: Logistic model, dairy cows, fatty acids, reproduction.

Sisukord

Sissejuhatus	4
1 Rasvhapped lehma piimas	6
2 Logistilise regressioonanalüüsi mudel	8
2.1 Fikseeritud faktoritega logistiline mudel	8
2.2 Juhuslike faktoritega logistiline mudel	11
2.3 Standardiseerimine	12
2.4 Mudeli karakteristikud	13
2.4.1 Dispersiooni inflatsioonitegur	13
2.4.2 ROC-kõver	14
2.4.3 Akaike informatsioonikriteerium	16
3 Rasvhapete analüüs	18
3.1 Andmestik	18
3.2 Esmane kirjeldav analüüs	19
3.2.1 Farmide võrdlus	19
3.2.2 Laktatsioonide võrdlus	23
3.3 Lehma tiinestuvuse hindamine rasvhapete abil	27
Kokkuvõte	35
Kasutatud allikad	36

Sissejuhatus

Eestis püsib üks piimalehm karjas keskmiselt 2,5 laktatsiooni ehk poegimisjärgset perioodi, mil lehm toodab piima. Kuna karjapidajale kasumlikuks muutub lehm alles teise laktatsiooni keskpaigas, on pidev uute loomade lisandumine ehk karja taastootmine majanduslikult võtmetähtsusega. Karja taastootmiseks peavad lehmad poegima, viimase eelduseks on omakorda see, et lehmad jäävad tiineks. Lehmade seemendamiseks kasutatakse tänapäeval üldjuhul seemendustehnikut, kes sooritab seemendusi 21-päevaste intervallidega kuni lehma õnnestunud tiinestumiseni. Igale loomale sobivaima seemendusaja valimine ja selle tulemusena potentsiaalselt toimuva tiinestumise prognoosimine võimaldaks muuta seemendamise protsessi efektiivsemaks ja langetada sellega kaasnevaid kulusid.

Peaaegu kõigi Eesti piimalehmade piimatoodangu ja -kvaliteedi kohta kogutakse infot kord kuus toimuvatel kontroll-lüpsidel, kus registreeritakse lüpsitud piima kogus ja saadetakse piimaproov laborisse edasisteks analüüsideks. Selle kõigega tegeleb Eesti Põllumajandusloomade Jõudluskontrolli AS (EPJ). Alates aastast 2022 on EPJ hakanud osades kontroll-lüpsi piimaproovides määrama ka erinevate rasvhapete sisaldust. Rasvhappelist koostist määratakse hetkeseisuga 34 Eesti farmis. Bakalaureusetöö eesmärk on välja selgitada, kas kontroll-lüpsil määratud piima rasvhappeline sisaldus annab piisavalt informatsiooni lehma tiinestumise kohta ning kui täpset prognoosi rasvhappe koostise teadmine võimaldab.

Bakalaureusetöö koosneb kolmest suurest peatükist. Esimene peatükk annab ülevaate rasvhapete määramise tehnoloogiast, rasvhapete peamisest liigitusest ning toob välja faktorid, mis mõjutavad rasvhapete kogust lehmapiimas. Teine peatükk kirjeldab analüüsis kasutatud juhuslike efektidega logistilise

regressiooni mudelit ja selle tõlgendamist ning tutvustab mudeli sobivuse hindamiseks kasutatavaid karakteristikuid. Kolmandas peatükis on kirjeldatud EPJ-i andmetel läbi viidud analüüsi tulemusi.

1 Rasvhapped lehma piimas

Kuigi piima rasva ja valgu sisaldust on määratud juba enam kui sajand, on piima rasvhappelise koostise rutiinne tuvastamine võrdlemisi uus praktika. Tõuke rasvhappelise koostise rutiinsele määramisele andis Fourier' teisendusega infrapunaspektroskoopia (FTIR) areng, mis parandas märkimisväärselt prognoosimisvõimekust. Võrdluses alternatiivse meetodiga, gaasikromatograafiaga, on infrapunaspektroskoopia kiirem, odavam ja ei nõua erialaspetsialistide talitlust. Fourier' teisendusega infrapunaspektroskoope kasutatakse piima kvaliteedinäitajate määramiseks nii üksikute lehmade kui ka segupiima piimaproovides. Eelkõige kasutatakse FTIR-i ametlikes piimakontrolli süsteemides valgu, kaseiini, rasva, laktoosi ja karbamiidi sisalduse prognoosimiseks. Lisaks nendele traditsioonilistele näitajatele on FTIR-i kasutusvaldkond pidevalt laienenud ja laienemas ka teiste piima koostiskomponentide ja omaduste prognoosimiseks, nagu piima rasvhappeline koostis, piimavalkude koostis, piima laapumisomadused, ketokehade sisaldus jmt näitajad. (De Marchi *et al.*, [2014](#))

Piimaanalüsaatoriga määratud piimaproovide rasvhappelise profiili kasutamise praktilisteks rakendusteks võiksid olla piimatöötlemise optimeerimine ja/või piimalehmade jõudluse parandamine parema söötmise ja juhtimise kaudu ja negatiivse energiabilansiga seonduvate terviseprobleemide varajane prognoosimine (Schwarz, Bak ja Hansen, [2022](#); Churakov *et al.*, [2021](#)).

De novo sünteesitakse udaras rasvhapped, mille süsinikahela pikkus on 4 kuni 16 süsinikku. Nende rasvhapete sünteesi prekursoriteks on vatsas moodustuvad lenduvad rasvhapped, peamiselt atsetaat ja butüraat. *De novo* rasvhapped on hea indikaator lehma vatsa tervise hindamiseks, optimaalne sisaldus viitab vatsa heale mikroobikooslusele ja edukale süsivesikute fermentatsioonile.

nile. *De novo* rasvhapete sisalduse langus võib aga viidata ratsiooni tasakaalustamatusele erinevate toitainete osas (lõhustuv tärklis ja kiudained) või negatiivse energiabilansiga kaasnevale valmis rasvhapete osakaalu suurenemisele. (McGuire ja Bauman, 2002; Pentjärv ja Henno, 2025)

Valmis rasvhapped pärinevad söödalipiididest ja keharasvade mobilisatsioonist ning liiguvad udarasse vereringe kaudu, nende süsinikahelas on 16 või rohkem süsinikku (McGuire ja Bauman, 2002). Laktatsiooni algul on lehma suurenenud energiavajadusest tulenevalt märgata selles proportsioonis muutust. Kasvanud energiavajaduse rahuldamiseks mobiliseeritakse rohkem kehavarusid ning halvimal juhul satub lehm negatiivsesse energiabilanssi ehk energiakulu ületab söödast saadud energiat. Lisaks negatiivse energiabilansiga kaasnevale valmis rasvhapete sisalduse suurenemisele piimarasvas võivad samasuunalisi muutusi põhjustada ka rasvarikaste söötade sisalduse suurenemine söödaratsioonis ning ka lehmade söömuse suurenemine. (Pentjärv ja Henno, 2025)

Eriliigilised rasvhapped võivad olla nii *de novo* udaras sünteesitud kui ka vereringest otse võetavad rasvhapped ehk valmis rasvhapped. Nende rasvhapete sisalduse suurenemine piimarasvas kaasneb peamiselt rasvhapete *de novo* sünteesi osatähtsuse suurenemisega ning ka palmiõli söötmisega. Lehma seedimist mõjutavatel faktoritel on niisiis otsene mõju ka eriliigiliste rasvhapete sünteesile. (Pentjärv ja Henno, 2025)

2 Logistilise regressioonanalüüsi mudel

Logistilise regressiooni mudel on statistiline meetod, mida kasutatakse binaarse (jah/ei) tunnuse modelleerimiseks. Logistilise mudeli eesmärk on sõltumatute muutujate põhjal hinnata mingi sündmuse toimumise tõenäosust. Selles peatükis on mudeli kirjeldamiseks kasutatud Alan Agresti raamatut „Categorical Data Analysis“ (Agresti, 2002) ja Charles E. McCullochi ning Shayle R. Searle'i raamatut „Generalized, Linear and Mixed Models“ (McCulloh ja Searle, 2001).

2.1 Fikseeritud faktoritega logistiline mudel

Olgu Y binaarne uuritav tunnus, X sõltumatu argumenttunnus ja sündmuse toimumise tõenäosus

$$\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x).$$

Lihtne logistiline mudel avaldub seljuhul kujul

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}.$$

Sündmuse toimumise tõenäosuse $\pi(x)$ ja mittetoimumise tõenäosuse $1 - \pi(x)$ suhet

$$\frac{\pi(x)}{1 - \pi(x)}$$

nimetatakse sündmuse toimumise šansiks. Sündmuse toimumise šansi abil saab logistilise mudeli esitada lineaarse ja kogu arvteljel pideva funktsioonina

$$\text{logit}[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x, \quad (1)$$

kus α on vabaliige, x argumenttunnuse X väärtus ning β argumenttunuse mõju tugevust ja suunda väljendav regressioonikordaja. Regressioonikordaja ees olev märk näitab seda, kas sündmuse toimumise tõenäosus x -i kasvades kasvab või kahaneb.

Võrrandi (1) mõlemad pooled e astmesse tõstes avaldub mudel šansi kaudu eksponentsiaalsel kujul

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x). \quad (2)$$

Esitusest (2) tulenevalt näitab $\exp(\alpha)$ sündmuse toimumise šanssi $X = 0$ korral

$$\frac{\pi(X = 0)}{1 - \pi(X = 0)} = \exp(\alpha + \beta \cdot 0) = \exp(\alpha),$$

ja $\exp(\beta)$ näitab sündmuse toimumise šansi multiplikatiivset kasvu argumenttunnuse X väärtuse üheühikulisel suurenemisel

$$\frac{\pi(x+1)}{1 - \pi(x+1)} = \exp(\alpha + \beta(x+1)) = \exp(\alpha + \beta x) \cdot \exp(\beta) = \frac{\pi(x)}{1 - \pi(x)} \cdot \exp(\beta),$$

siit avaldub omakorda

$$\exp(\beta) = \frac{\frac{\pi(x+1)}{1 - \pi(x+1)}}{\frac{\pi(x)}{1 - \pi(x)}}.$$

Viimast suhet nimetatakse šansside suhteks (ingl. *odds ratio*, OR).

Logistilist regressiooni on võimalik rakendada ka mitme argumenttunnuse korral. Olgu taaskord $\pi(\mathbf{x}) = P(Y = 1)$ sündmuse toimumise tõenäosus ning p muutujaga argumenttunnuste väärtuste komplekt $\mathbf{x} = (x_1, x_2, \dots, x_p)$, siis mudel avaldub kujul

$$\text{logit}[\pi(\mathbf{x})] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

Tõenäosuse $\pi(\mathbf{x})$ kaudu esitub mudel niisiis kujul

$$\pi(\mathbf{x}) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)},$$

kus parameeter β_i , $i \in \{1, \dots, p\}$, on tunnuse X_i mõjule vastav regressioonikordaja ja α on vabaliige.

Kui argumenttunnus X on diskreetne faktortunnus tasemetega $x_1, \dots, x_k$, esitub mudel sarnaselt mitmesele logistilisele regressioonile:

$$\text{logit}[\pi(x)] = \alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k),$$

kus

$$\mathbb{I}(X = x_i) = \begin{cases} 1 & , \text{ kui } X = x_i \\ 0 & , \text{ kui } X \neq x_i \end{cases}, i = 1, \dots, k,$$

on tunnuse X i -nda taseme x_i indikaatorfunktsioon ja β_i on selle taseme mõju. Diskreetse faktortunnuse puhul võetakse parameetrite ühese hinnangu saamiseks enamasti üks tase (näiteks x_1) baastasemeks ja sellele vastav mõju võrdseks nulliga ($\beta_1 = 0$). Sellest tulenevalt iseloomustab mudeli vabaliige sündmuse toimumise šanssi baastasemel:

$$\frac{\pi(X = x_1)}{1 - \pi(X = x_1)} = \exp(\alpha + \beta_1 \mathbb{I}(X = x_1)) = \exp(\alpha).$$

Ülejäänud tasemete mõjud näitavad erinevust baastasemest. Seejuures kujutab suhe

$$\frac{\exp[\alpha + \beta_i \mathbb{I}(X = x_i)]}{\exp[\alpha + \beta_1 \mathbb{I}(X = x_1)]} = \frac{\exp(\alpha) \cdot \exp(\beta_i)}{\exp(\alpha)} = \exp(\beta_i)$$

enesest šansside suhet

$$OR = \frac{\frac{\pi(x_i)}{1 - \pi(x_i)}}{\frac{\pi(x_1)}{1 - \pi(x_1)}}.$$

Niisiis näitab eksponent i -nda grupi mõjust, kui mitu korda erineb sündmuse toimumise šanss grupis i sündmuse toimumise šansist baasgrupis.

2.2 Juhuslike faktoritega logistiline mudel

Kordusmõõtmiste korral moodustab iga vaatlusobjekt, millel mõõtmisi teostatakse klatri. Üldjuhul on klastrisisene varieeruvus väiksem kui klastrite vaheline varieeruvus ehk teisisõnu on klastris üksteisele sarnased vaatlused. Juhusliku faktori tasemetele vastavate klastrite vaheline erinevus omistatakse mittemõõdetavatele teguritele, mille mõju varieerub klastrites juhuslikult. Sellest tulenevalt kutsutaksegi neid juhuslikeks efektideks. Juhuslike faktoritega mudel hindab tinglikud parameetrite väärtused, kus tingimuseks on teatud klatri realiseerunud juhuslik efekt.

Olgu $\mathbf{x} = (x_1, \dots, x_p)$ argumenttunnuste $X_1, \dots, X_p$ realiseerunud väärtuste komplekt ning $Z_1, \dots, Z_t$ diskreetsed juhuslikud faktortunnused, mille korral on realiseerunud tasemed $z_1, \dots, z_t$ ja nende tasemete mõjud $u_1, \dots, u_t$,

kus $u_j \sim N(0, \sigma_z^2)$, $j = 1, \dots, t$. Juhuslike efektide realiseerunud väärtuste tingimusel avaldub mudel kujul

$$\text{logit}(\pi(\mathbf{x}|z)) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p + u_1 \mathbb{I}(Z_1 = z_1) + \dots + u_t \mathbb{I}(Z_t = z_t),$$

siit avaldub huvipakkuva sündmuse tinglik tõenäosus

$$\pi(\mathbf{x}|z) = \frac{\exp(\alpha + \beta_1 x_1 + \dots + \beta_p x_p + u_1 \mathbb{I}(Z_1 = z_1) + \dots + u_t \mathbb{I}(Z_t = z_t))}{1 + \exp(\alpha + \beta_1 x_1 + \dots + \beta_p x_p + u_1 \mathbb{I}(Z_1 = z_1) + \dots + u_t \mathbb{I}(Z_t = z_t))}.$$

2.3 Standardiseerimine

Juhuslike efektidega logistilise regressiooni mudeli hindamine on arvutusmahukas protsess, olukorra parandamiseks kasutatakse tihti argumenttunnuste standardiseerimist. Standardiseerimine võimaldab pidevad arvulised tunnused viia ühisele skaalale, mis lihtsustab parameetrite hindamist. Enim kasutatud standardiseerimise funktsioon on Z-skoor:

$$Z_{\text{skoor}} = \frac{X - \mu}{\sigma},$$

kus X on mudelis kasutatav argumenttunnus, μ ja σ on selle tunnuse keskvärtus ning standardhälve. Sisuliselt näitab Z-skoor, mitme standardhälbe võrra erineb vaatlus keskmisest. (Wang *et al.*, 2023)

2.4 Mudeli karakteristikud

2.4.1 Dispersiooni inflatsioonitegur

Selle alapeatüki kirjutamisel on kasutatud Roman Salmeron Gomezi jt teadusartiklit „The VIF and MSE in Raise Regression“ (Gomez *et al.*, 2020).

Dispersiooni inflatsioonitegur (ingl. *Variance Inflation Factor*, VIF) on karakteristik, mida kasutatakse mudeli sõltumatute argumenttunnuste vahelise multikollineaarsuse tuvastamiseks. VIF arvutatakse igale mudeli argumenttunnusele. Olgu meil $X_1, \dots, X_p$ argumenttunnust, VIF_j leidmiseks hinnatakse j -ndale argumendile ülejäänud argumenttunnuste $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p$ põhjal lineaarse regressiooni mudel

$$X_j = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \dots + \alpha_p X_p,$$

kus α_1 on vabaliige ja α_k ($k \in \{2, \dots, j-1, j+1, \dots, p\}$) on regressioonikordajad. Selle mudeli determinatsioonikordaja $R_j^2 \in [0, 1]$ abil avaldub j -nda argumenttunnuse dispersiooni inflatsioonitegur kujul

$$VIF_j = \frac{1}{1 - R_j^2}.$$

Juhul, kui argumendil X_j puudub ülejäänud argumenttunnustega lineaarne seos on determinatsioonikordaja R_j^2 väärtus 0 ehk $VIF_j = 1$. Lineaarse seose kasvades kasvab determinatsioonikordaja väärtus ning seeläbi ka VIF-i väärtus. Praktikas loetakse multikollineaarsuse olemasolu kinnitatuks, kui VIF-i väärtus on suurem kui 10.

Segamudeli (juhuslike efektidega logistiline regressioon) korral toimub multikollineaarsuse tuvastamine fikseeritud efektide seas ehk juhuslikud efektid

jäetakse VIF-i arvutamisel kõrvale.

2.4.2 ROC-kõver

Selle alapeatüki kirjutamisel on kasutatud Jane V. Carteri jt teadusartiklit „ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves“ (Carter *et al.*, 2016).

Binaarse prognoositava tunnusega mudeli üks olulisim headusnäitaja on selle mudeli täpsus ehk, kui hästi suudab mudel eristada üht klassi teisest. Üks headuse analüüsimisel kasutatavaid meetodeid on ROC-kõver. Selle kõvera tõlgendamiseks tuleb esmalt mõista mudeli klassifitseerimisprotsessi. Olgu binaarse tunnuse väärtus 1 positiivne tulemus ja väärtus 0 negatiivne tulemus. Selle tunnuse klassifitseerimise tulemus on võimalik liigitada nelja kategooriasse:

1. Tõeselt positiivne (ingl. *true positive*, TP) - mudeli klassifitseeritud positiivne väärtus on ka tegelikkuses positiivne väärtus
2. Valepositiivne (ingl. *false positive*, FP) - mudeli klassifitseeritud positiivne väärtus on tegelikkuses negatiivne väärtus
3. Tõeselt negatiivne (ingl. *true negative*, TN) - mudeli klassifitseeritud negatiivne väärtus on ka tegelikkuses negatiivne väärtus
4. Valenegatiivne (ingl. *false negative*, FN) - mudeli klassifitseeritud negatiivne väärtus on tegelikkuses positiivne väärtus

Mudeli täpsuse hindamisel kasutatakse nimetatud tulemuste osakaalude leidmist, neist kõige populaarsemad näitajad on tundlikkus ja spetsiifilisus.

Tundlikkus kirjeldab edukalt klassifitseeritud positiivsete tulemuste osakaalu kõikide positiivsete vaatluste seas. Kasutades eelnevalt defineeritud tähistusi saab tundlikkuse avaldada kujul

$$\text{Tundlikkus} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

Spetsiifilisus näitab aga vastupidiselt korrektselt klassifitseeritud negatiivsete tulemuste osakaalu kõikide negatiivsete vaatluste seas. Defineeritud tähistuste abil saab spetsiifilisuse kirja panna kui

$$\text{Spetsiifilisus} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Spetsiifilisuse asemel on laialdasemat kasutust leidnud näitaja $1 - \text{Spetsiifilisus}$. On lihtne veenduda, et sellise näitaja puhul on tegu valepositiivsete osakaaluga kõigi negatiivsete vaatluste seas:

$$1 - \text{Spetsiifilisus} = 1 - \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{TN} + \text{FP}}{\text{TN} + \text{FP}} - \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{FP}}{\text{TN} + \text{FP}}.$$

Kasutades klassifitseerimisel logistilist regressiooni tuleb valida otsustuslävend (üldjuhul $p = 0,5$), mille ületamisel liigitatakse prognoos positiivseks tulemuseks. **ROC-kõver** on mudeli täpsuse graafiline esitus, kus on kujutatud erinevatele otsustuslävenditele vastavaid tundlikkuse ja spetsiifilisuse paare, mis on omavahel graafiliselt ühendatud. Enamasti on tundlikkust kujutatud y-teljel ning x-teljel on kujutatud $1 - \text{Spetsiifilisus}$ ehk valepositiivsete osakaal.

Üks oluline mudeli headusnäitaja on ROC-kõvera alune pindala (AUC). Kõ-

veraalse pindala väärtus 0,5 viitab sellele, et mudeli prognoos on täiesti juhuslik ja argumenttunnused ei oma mingit kirjeldusvõimet prognoostunnuse osas. AUC väärtuse 1 korral oleks tegu ideaalse mudeliga. Üldpildis liigitatakse prognoosimiseks mõeldud mudelid AUC põhjal järgmiselt:

- $0,5 < \text{AUC} \leq 0,7$ - mudel ei sobi prognoosimiseks,
- $0,7 < \text{AUC} < 0,8$ - mudel on aktsepteeritava prognoosimisvõimega,
- $0,8 \leq \text{AUC} < 0,9$ - mudel on tugeva prognoosimisvõimega,
- $0,9 \leq \text{AUC}$ - mudel on suurepärase prognoosimisvõimega.

2.4.3 Akaike informatsioonikriteerium

Selle alapeatüki kirjutamisel on kasutatud David W. Hosmer juuniori jt raamatut „Applied Logistic Regression“ (Hosmer Jr., Lemeshow ja Sturdivant, 2013).

Samal andmestikul aga erinevate parameetrite arvuga hinnatud mudelite võrdlemiseks kasutatakse Akaike informatsioonikriteeriumit (AIC), mille arvutamine tugineb tõepärafunktsioonile. Olgu meil n vaatluspaari $(\mathbf{x}_i, y_i)$, kus y_i on uuritav binaarne tunnus, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ on i -nda vaatluse argumenttunnuste väärtuste komplekt ja sündmuse $y_i = 1$ toimumise tõenäosus on

$$\pi(\mathbf{x}_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})},$$

ehk sündumuse toimumise tõenäosus vastab mitmesele logistilisele regressioonile. Kasutades ka vastandsündmuse tõenäosust saab vaatluspaari $(\mathbf{x}_i, y_i)$ realiseerunud sündmuse tõenäosuse üldkujul kirja panna kui

$$\pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}.$$

Logistilise regressiooni tõepärafunktsioon $l(\boldsymbol{\beta})$ on eelneva avaldise korrutis üle kõikide vaatluste

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}.$$

Logaritmiline tõepära on naturaalogaritm eelnevast funktsioonist:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_i^n [y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]].$$

Logaritmilist tõepära regressioonikordajate $\beta_1, \dots, \beta_p$ suhtes maksimeerides saab suurima tõepära hinnangu $L(\hat{\boldsymbol{\beta}})$

$$L(\hat{\boldsymbol{\beta}}) = \max_{\beta_1, \dots, \beta_p} L(\boldsymbol{\beta}).$$

Akaike informatsioonikriteerium kasutab suurima tõepära hinnangut võrdlemaks mudeli sobivust andmetele ja selle mudeli keerukust:

$$\text{AIC} = -2L(\hat{\boldsymbol{\beta}}) + 2(p + 1),$$

kus p on argumenttunnuste arv mudelis. Mudelite võrdlemisel eelistatakse mudelit, millel on madalam AIC väärtus.

3 Rasvhapete analüüs

3.1 Andmestik

Analüüsis kasutatud andmed on pärit Eesti Põllumajandusloomade Jõudluskontrolli AS-i andmebaasist. Analüüsiks on kasutatud rakendustarkvara R.

Andmestik koosneb vahemikus 01.02.2022 kuni 18.12.2025 34 Eesti farmis kogutud eesti holsteini, eesti punase veisetõu ja eesti maakarja lehmade kontroll-lüpside tulemustest. Kokku on selles ajavahemikus teostatud ning analüüsitud 115 658 kontroll-lüpsi, pärast puuduvate või vigaselt sisestatud andmeridade eemaldamist jääb neist alles 98 168 vaatlust. Mudeli ja analüüsi lihtsustamise eesmärgil on oma vähese arvukuse tõttu (4951 vaatlust) eesti punase veisetõu ja eesti maakarja loomade mõõtmistulemused andmestikust eemaldatud. Niisiis keskendub analüüs eesti holsteini tõugu lehmade piima rasvhappelisele profiilile. Kokku on 11 323 holsteini lehmale määratud 93 217 kontroll-lüpsi tulemust. Uuritavate rasvhapete kogused on edaspidi esitatud ühikus grammi 100 milliliitri piima kohta.

Andmestikus on iga lehma kohta tunnus *Tiinnostav*, mille väärtus on 1, kui lehm jäi esimesest seemendusest tiineks ja 0, kui ei jäänud. Huvipakkuvaid argumenttunnuseid on viis:

1. *de novo* rasvhapete kogus kontroll-lüpsil
2. eriliigiliste rasvhapete kogus kontroll-lüpsil
3. valmis rasvhapet kogus kontroll-lüpsil
4. laktatsioonipäev ehk mitmendal laktatsioonipäeval kontroll-lüps tehti

5. laktatsioon, nelja tasemega faktortunnus, mis jagab lehmad 1., 2., 3. ning 4. ja enama laktatsiooni lehmadeks

Lisaks on teada iga lehma unikaalne registrikood, farmide ehk omanike koodid ning seemendustehnikute koodid.

3.2 Esmane kirjeldav analüüs

Rasvhapete dünaamika uurimiseks on 365-päevane vaatlusperiood jagatud 5-päevasteks intervallideks, igas sellises intervallis on leitud huvipakkuva grupi keskmine rasvhappe sisaldus. Laktatsiooni perioodi dünaamika visualiseerimiseks on saadud intervallide keskmisi kasutatud LOESS-i regressioonimudeli hindamisel, kus silumisparameeter on vaikimisi valitud 0,3.

3.2.1 Farmide võrdlus

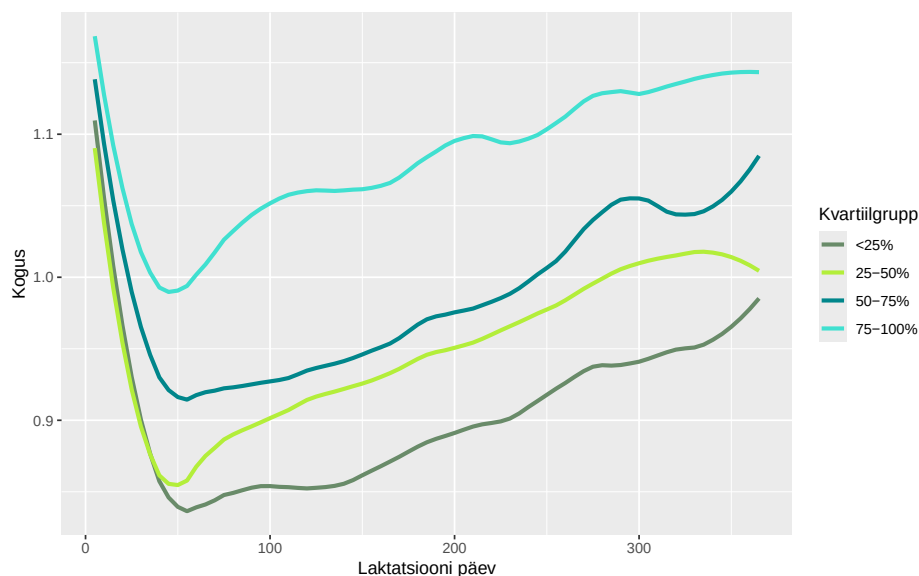
Farmide dünaamika kõverate võrdlemiseks LOESS-i regressiooniga on esmalt leitud iga farmi rasvhappe keskmine tase ning seejärel farmide keskmiste tasemete kvartiilid. Tabelis 1 toodud kvartiilide põhjal on andmestikku kuulunud 34 farmi jagatud nelja kvartiilgruppi. Esimese kvartiilgrupi moodustavad need farmid, kelle keskmine rasvhappe tase jääb alla alumise kvartiili. Järgnevad kaks kvartiilgruppi on kahe järjestikuse kvartiili vahemikud. Viimasesse ehk neljandasse gruppi kuuluvad farmid, kelle keskmine rasvhappe tase on kõrgem ülemisest kvartiilist. Järgnevalt on analüüsitud rasvhapete dünaamikat kvartiilgruppides.

Tabel 1. Farmide rasvhapete arvkarakteristikud (g/100 ml piima kohta).

Rasvhappe klass	Miinumum	Alumine kvartiil	Mediaan	Ülemine kvartiil	Maksimum
<i>De novo</i>	0.86	0.93	0.96	1.02	1.27
Valmis	1.30	1.42	1.51	1.56	1.70
Eriliigilised	1.02	1.10	1.12	1.20	1.58

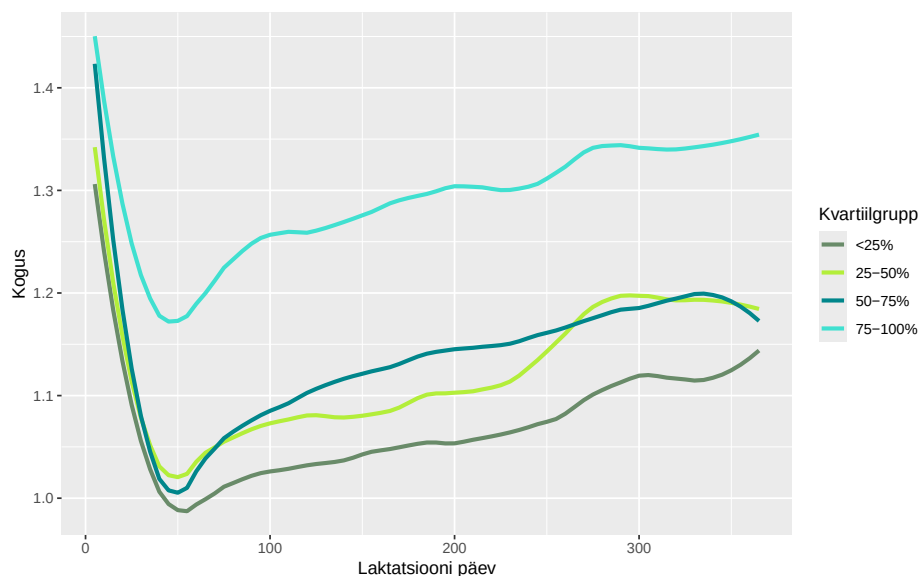
De novo rasvhapete kogus on laktatsiooni algul farmide kvartiilgruppidel võrdlemisi sarnane, püsides 1,1 ühiku ligidal (Joonis 1). Järgneva pooleteist kuu jooksul on igas grupis märgata järsku langust. Sarnast trendi järgivad esimene ja teine kvartiilgrupp, mis saavutavad oma miinimumi, ligikaudu 0,85 ühikut, 50 päeva kandis. Kolmandas ja neljandas grupis on esmane langus laugem ning miinimumi väärtus kõrgem, kolmandal grupil napilt alla 0,925 ühiku ja neljandal grupil alla 1 ühiku. Peale miinimumi saavutamist hakkab rasvhappe tase ühtlaselt taastuma. Suurima, enam kui 0,05-ühikulise, hüppe teeb miinimumile järgneva 50 päeva jooksul neljas grupp. Neljas kvartiilgrupp on ühtlasi ainuke, mis on 365 päeva möödudes taastunud enam-vähem algsele tasemele. Ülejäänud grupid järgivad võrdlemisi lineaarset kasvavat trendi, välja arvatud kolmanda grupi 250 ja 300 päeva vaheline 0,05 ühiku suurune spurt. Vaatlusperioodi lõpuks on kolme madalama kvartiilgrupi *de novo* rasvhappe tase taastunud enam kui 88% algsest tasemest. Üldjoontes võib öelda, et gruppide *de novo* rasvhapete dünaamika on sarnane, ent selgelt joonistub välja erinevus koguses.

Joonisel 2 näeme, et eriliigiliste rasvhapete keskmine on laktatsiooni algul kuni 0,3 ühiku võrra kõrgem *de novo* rasvhapete omast, olles esimese ja teise grupi puhul 1,325 kandis ning kahe viimase grupi puhul 1,4 ja 1,45 vahel. Sarnaselt *de novo* rasvhapetele toimub esimese 50 päeva jooksul eriliigiliste rasvhapete koguses järsk langus. Seekord käituvad aga esimesed kolm grup-



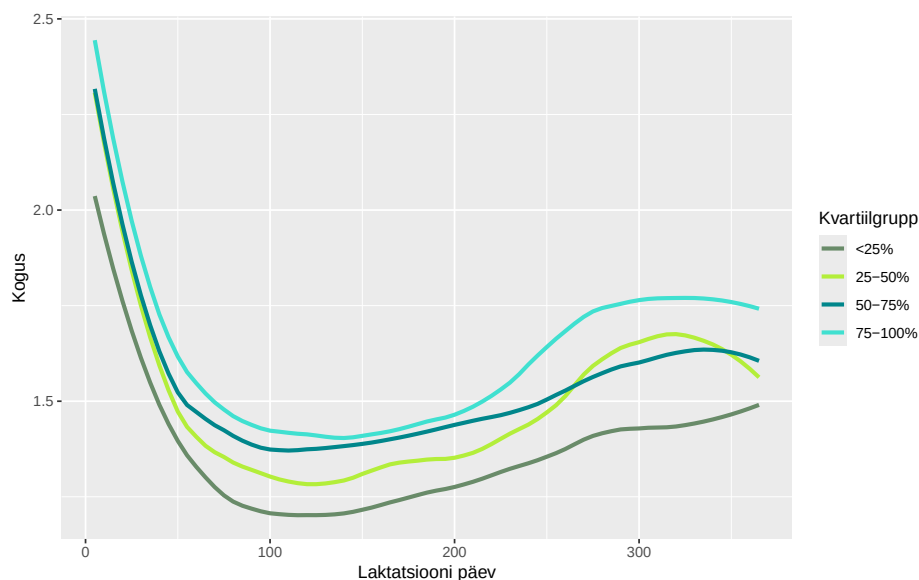
Joonis 1. *De novo* rasvhapete dünaamika farmides.

pi ühe trendi kohaselt, esmane erinevus on miinimumpunktis, kus esimene grupp saavutab teisest kahest natuke madalama väärtuse, langedes alla 1 ühiku. Neljanda grupi langus on teistest laugem, jäädes pidama 1,2 ühiku piiril. Kõige kiirem on miinimumist taastumine neljandal grupil, ent ka kaks keskmist gruppi demonstreerivad kiiret tõusu. Esimese kvartiilgrupi taastumine on tagasihoidlikum, suurem hüpe toimub 250 ja 300 päeva vahel ning aastaajase vaatlusperioodi lõpuks on selle grupi eriliigiliste rasvhapete kogus 87,6% algsest tasemest. Kaks keskmist gruppi järgivad sarnast kasvavat lineaarset trendi, ent kolmas grupp teeb vaatlusperioodi lõpus ootamatu languse. Seetõttu on kolmanda grupi taastumine madalaim, 82,4% algsest tasemest. Teine ja neljas grupp saavutavad vastavalt 88,2% ja 93,4% enda algsest tasemest. Üldpildis on dünaamika gruppides taaskord sarnane, väikese erandiga kolmandas grupis. Koguste drastilisem erinevus ilmneb neljanda grupi ja ülejäänud gruppide vahel.



Joonis 2. Eriliigiliste rasvhapete dünaamika farmides.

Joonisel 3 kujutatud valmis rasvhapete dünaamikas on märgata eelnevatest rasvhapetest erinevat mustrit. Taaskord on kogus suurim laktatsiooni algul, ent järgnev langus on iga grupi korral märgatavalt laugem. Miinimumid saavutatakse 100 ja 150 päeva vahel ning on proportsionaalsed gruppidega, jäädes alla 1,5 ühiku. Miinimumist taastumine on samuti tagasihoidlikum ning ei järgi teise ja viimase kvartiilgrupi puhul lineaarset trendi, miinimumijärgne maksimum saavutatakse vahetult peale 300-ndat päeva. Vaatlusperioodi lõpus toimuva väikse langusega jäävad gruppide kogused 67% ja 74% vahele nende algsest tasemest, seejuures parim taastumine on esimesel kvartiilgrupil. Esimene grupp on ühtlasi ainuke, kus ei toimu vaatlusperioodi lõpus langust ning saavutatakse miinimumijärgne maksimum. Kõige sügavam langus on perioodi lõpus teisel grupil, mistõttu on ka taastumine kõige kehvem, 67,7% algtasemest.



Joonis 3. Valmis rasvhapete dünaamika farmides.

3.2.2 Laktatsioonide võrdlus

Tabelis 2 on näha kontroll-lüpside jaotumist laktatsioonide seas. Peaaegu poole andmestikust moodustavad esimese laktatsiooni holsteini lehmade kontroll-lüpsi tulemused. Pea kaks korda väiksem on teise laktatsiooni lehmade vaatluste osakaal, moodustades veidi enam kui veerand mõõdetud tulemustest. Vaatluste osakaal jätkab kahanemist kolmanda ning enam kui kolmanda laktatsiooni lehmade seas, viimasesse neist kuulub iga kümnes mõõtmistulemus.

Tabel 2. Vaatluste jaotus laktatsioonide seas.

Laktatsioon	1.	2.	3.	> 3	Kokku
Sagedus	46 064	24 430	12 540	10 183	93 217
Protsent	49,4%	26,2%	13,5%	10,9%	100%

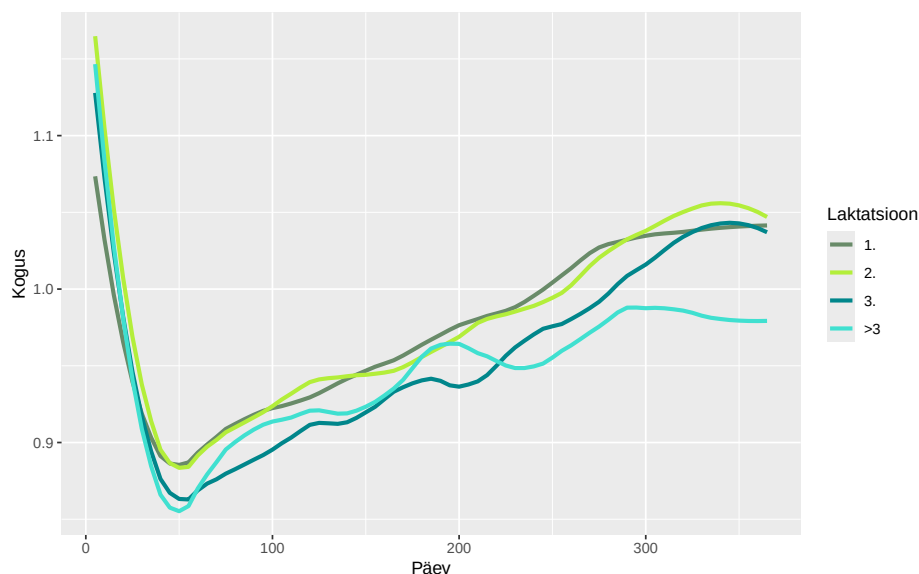
Uurides omakorda edukate esmaseemenduste osakaalu laktatsioonidel (Tabel 3) on selgelt näha, et lehma võime tiinestuda esimesel seemendusel kahaneb iga järgneva poegimisega. Kui esimesel laktatsioonil olevatel lehma-

del on tõenäosus esimesel seemendusel tiinestuda veel napilt üle 50%, siis järgnevatel laktatsioonidel on see juba alla 50%. Edukate esmaseemenduste osakaal langeb stabiilselt, kukkudes esimese laktatsiooni lehmade 51,2%-lt teise laktatsiooni lehmade 44,5%-le ning omakorda kolmanda laktatsiooni lehmadel 42,8%-le. Erandina on märgata 3-protsendipunktilist kasvu enam kui kolmanda laktatsiooni lehmade seas. Võttes arvesse, et keskmiselt on eesti holsteini piimalehmal 2,5 laktatsiooni, võib eelnevat anomaaliat seostada üksikute erakordselt hea viljakusega lehmadega selles laktatsiooni grupis.

Tabel 3. Tiinestavate esmaseemenduste osakaal laktatsioonidel.

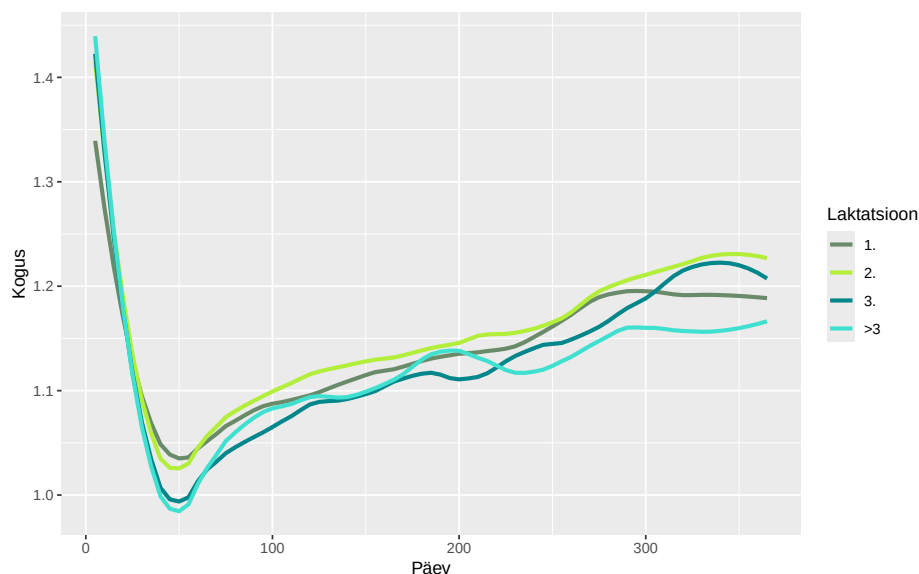
Laktatsioon/ Tiinestav esmaseemendus	1.	2.	3.	> 3
Jah	51,2%	44,5%	42,8%	45,9%
Ei	48,8%	55,5%	57,2%	54,1%

Võrreldes Joonist 4 Joonisega 1 on näha, et laktatsioonide *de novo* rasvhapete dünaamikas ei joonistu nii drastilist erinevust välja kui farmide võrdluses. Esmane tase laktatsiooni algul on gruppidel sarnane, enim erineb esimese laktatsiooni grupi kogus, mis on ainsana alla 1,1 ühiku. Taaskord on märgata koguse järsku langust ning miinimumi saavutamist 50-ndal päeval. Kuigi esimese ja teise laktatsiooni miinimumi ning kolmanda laktatsiooni miinimumi vahel on väikene erinevus, taastuvad nad seejärel lineaarses trendis üpriski võrdselt enam kui 89-protsendilisele tasemele algsest kogusest. Ainus erand on enam kui kolmanda laktatsiooni lehmade dünaamika, kus suuremate kõikumistega taastumine stabiliseerub peale 300-ndat päeva ning on aastase vaatlusperioodi lõpuks 85,4% algsest tasemest. Kõige jõudsam on esimese kvartilgrupi taastumine, jäädes 97,0%-ga napilt alla vaatlusperioodi algsele tasemele.



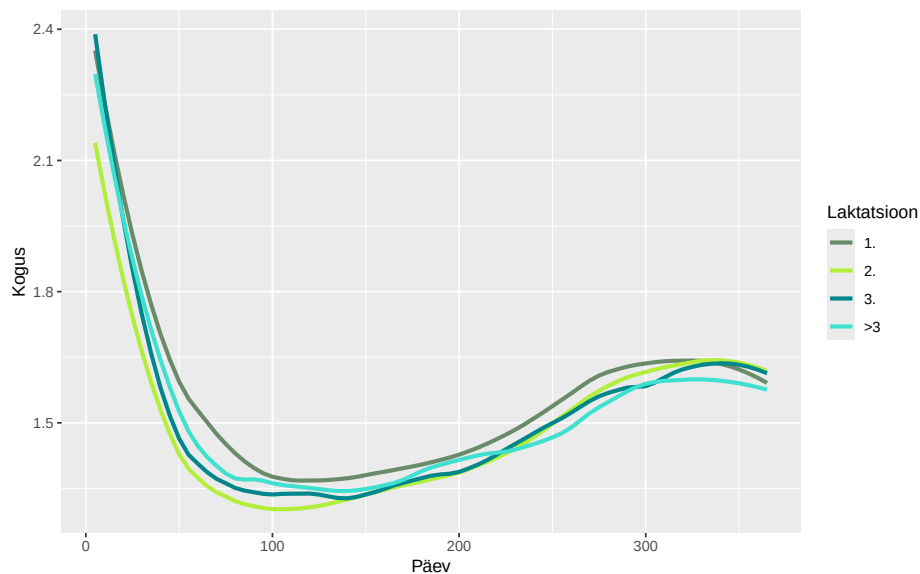
Joonis 4. *De novo* rasvhapete dünaamika laktatsioonidel.

Eriliigiliste rasvhapete kogus on laktatsiooni algul kõige madalam esimese laktatsiooni lehmadel, 1,34 ühikut, kuid sellele järgnev langus on sügavam kolmanda ja enama laktatsiooni lehmadel (Joonis 5). Jällegi on 50-ndal päeval saavutatav miinimum enam-vähem võrdne esimesel ja teisel laktatsioonil, olles kuni 0,05 ühiku võrra kõrgem ülejäänud laktatsioonidest. Miinimumist taastumine toimub taaskord lineaarses trendis, kusjuures enam kui kolmanda laktatsiooni lehmadel on see märgatavalt ebastabiilsem. Ka vaatlusperioodi lõpuks on vanimate lehmade rasvhappe tase kõige madalam ning kõige vähem taastunud, 81,0% algsest tasemest. Kõige kõrgem on eriliigiliste rasvhapete tase perioodi lõpus teise laktatsiooni loomadel, 1,23 ühikut, ent parim taastumine on hoopis esimesel laktatsioonil, 88,7% algtasemest võrreldes teise laktatsiooni 86,7%-ga.



Joonis 5. Eriliigiliste rasvhapete dünaamika laktatsioonidel.

Kui eelmiste rasvhapete gruppide korral leidsid selgeid eristumisi, siis valmis rasvhapete dünaamika (Joonis 6) liigub laktatsioonide lõikes sarnases taktis. Ainus suurem erinevus on rasvhapete algtasemes, mis on kõige madalam teise laktatsiooni loomadel, jäädes ainsana alla 2,25 piiri. Ülejäänud laktatsioonide algtase on 2,35 ühiku ümbruses. Taaskord järgneb igas grupis taseme langus, mis aga võrdluses eelnevate rasvhapetega on laugem. Vahemikus 100 ja 150 päeva saavutavad laktatsioonide kõverad oma miinimumväärtuse 1,35 ühiku ümbruses. Miinimumijärgne maksimum saavutatakse enne 350 laktatsioonipäeva ning jääb napilt alla 1,65 ühiku. Laktatsiooniperioodi lõpus on rasvhappe koguses minimaalne langus. Ka madalseisust taastumine on valmis rasvhapete puhul tagasihoidlikum, peale vaatlusperioodi lõpus toimuvat väikest langust jäävad tasemed 67% ligidale algtasemest. Parim on taastumine teise laktatsiooni loomadel, saavutades vaatlusperioodi lõpus 75,6% algsest kogusest.



Joonis 6. Valmis rasvhapete dünaamika laktatsioonidel.

3.3 Lehma tiinestuvuse hindamine rasvhapete abil

Algandmestikus on ühel real kontroll-lüpsi tulemused ning sellele lüpsile vahetult eelnenud seemenduste ja poegimise andmed. Rasvhapete abil lehma tiinestuvuse prognoosimiseks on aga vaja teada rasvhappe kogust enne seemendust. Niisiis on algandmestikust eraldatud kolm alamandmestikku:

- Kuni 30 päeva enne esimest seemendust toimunud kontroll-lüpsi tulemused. Selles andmestikus on 2869 unikaalsel lehmalt registreeritud 3592 kontroll-lüpsi.
- Kuni 45 päeva enne esimest seemendust toimunud kontroll-lüpsi tulemused. Selles andmestikus on 3349 unikaalsel lehmalt registreeritud 5309 kontroll-lüpsi.
- Kuni 60 päeva enne esimest seemendust toimunud kontroll-lüpsi tulemused. Selles andmestikus on 3690 unikaalsel lehmalt registreeritud 6732 kontroll-lüpsi.

Üldjuhul teostavad karjapidajad kontroll-lüpsse 20–30-päevastes intervallides. Niisiis on igas andmestikus lehmi, kellel on samale seemendusele eelnevalt registreeritud mitme kontroll-lüpsi andmed (andmestikus mitu erinevat vaatlust). Enim on mitme lüpsitulemusega lehmi 60-päevases andmestikus, 1885 lehma.

Rasvhapete ja tiinestumise seose uurimiseks on iga andmestiku korral hinnatud logit-seosefunktsiooniga üldistatud lineaarne segamudel ehk juhuslike efektidega logistiline regressioon, kus uuritav tunnus on esimesel seemendusel tiinestumine (jah/ei). Argumenttunnused on iga mudeli korral *De novo* rasvhapete kogus, *Eriliigiliste* rasvhapete kogus, *Valmis* rasvhapete kogus, *Laktatsioonipäev* ja *Laktatsioon*. Juhuslikud faktorid on *Lehm*, *Omanik* ja *Seemendustehnik*.

Esmased mudelid on hinnatud kahel juhul, esmalt kasutades kõikide argumenttunnuste peamõjusid ning siis lisades ka teistjärku koosmõjud rasvhapete ja laktatsioonide ning rasvhapete ja laktatsioonipäeva vahel. Tabelis 4 on aga näha, et koosmõjud ei anna ühegi andmestiku korral olulist lisainformatsiooni tiinestuvuse kirjeldamisel (AIC väärtus on koosmõjudega mudelitel kõrgem). Seetõttu on edaspidises mudeli analüüsis uuritud ainult peamõjusid.

Tabel 4. Muutus Akaike informatsioonikriteeriumis koosmõjude lisamisel.

Mudel/ Koosmõjudega	30-päevane	45-päevane	60-päevane
Ei	4911,7	6942,1	8488,7
Jah	4925,7	6958,2	8506,6
Muutus	+14,0	+16,1	+17,9

Enne peamõjudega mudelite võrdlust tuleb välja selgitada, kas nende mudelite argumenttunnuste vahel esineb multikollineaarsust. Selle tuvastamiseks on kasutatud dispersiooni inflatsioonitegurit ehk VIF-i. Tabelist 5 nähtub, et *de*

novo ja eriliigiliste rasvhapete tunnuste VIF väärtused on kõigis mudelites suhteliselt kõrged (≈ 7), mis viitab multikollineaarsusele. Arvestades nende kahe rasvhappe liigi bioloogilist seotust, on selline tulemus ootuspärane. Teiste tunnuste VIF väärtused jäävad seejuures madalaks ega viita olulisele multikollineaarsusele. Multikollineaarsuse vähendamiseks on suurima dispersiooni inflatsiooniteguriga tunnus *Eriliigilised rasvhapped* edasisest analüüsist kõrvale jäetud.

Tabel 5. Peamõjudega mudelite dispersiooni inflatsioonitegurid.

Mudel/ Argumenttunnus	30-päevane	45-päevane	60-päevane
<i>De novo</i> rasvhapped	7,62	6,96	6,70
Eriliigilised rasvhapped	7,80	7,31	7,22
Valmis rasvhapped	1,38	1,45	1,59
Laktatsioonipäev	1,05	1,12	1,21
Laktatsioon	1,05	1,05	1,04

Hinnates uued mudelid ilma tunnusetä *Eriliigilised rasvhapped* näeme, et enam pole põhjust multikollineaarsust kahtlustada, tunnuste VIF väärtused jäävad vahemikku $[1,03; 1,46]$ (Tabel 6). Niisiis on jätkatud sobiva mudeli hindamist nelja argumenttunnusega.

Tabel 6. Peamõjudega mudelite dispersiooni inflatsioonitegurid peale tunnuse *Eriliigilised rasvhapped* eemaldamist.

Mudel/ Argumenttunnus	30-päevane	45-päevane	60-päevane
<i>De novo</i> rasvhapped	1,28	1,23	1,22
Valmis rasvhapped	1,35	1,38	1,46
Laktatsioonipäev	1,04	1,12	1,20
Laktatsioon	1,05	1,04	1,03

Nelja tunnusega hinnatud mudelite parameetreid uurides on näha, et tunnuse *Laktatsioon* tasemete erinevus baastasemest ehk esimese laktatsiooni leh-

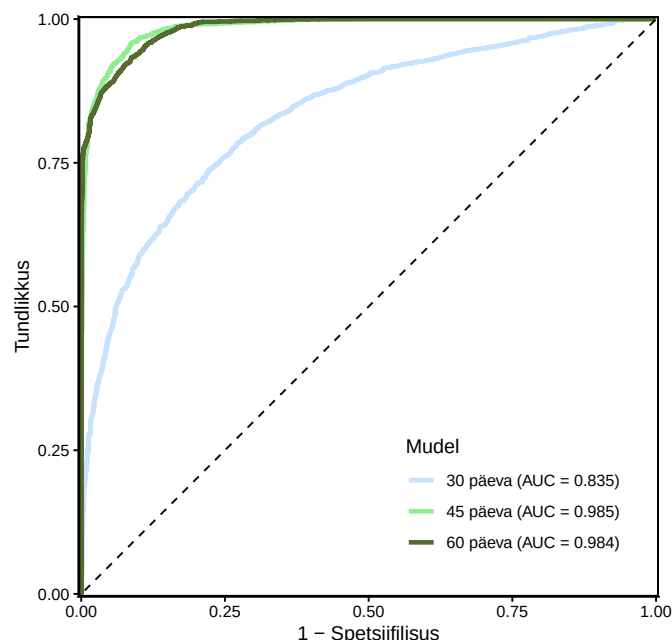
madest on iga mudeli korral statistiliselt oluline (Tabel 7). Võrreldes nende tasemete parameetrite väärtusi on aga näha, et erinevused pole väga suured, 60-päevase mudeli korral on iga taseme puhul hinnang ligikaudu $-0,41$. Hinnangute sarnasus viitab sellele, et teise laktatsiooni ja kolmanda ning enam kui kolmanda laktatsiooni loomade tõenäosus tiinestuda on sarnane. Tukey korrektsiooniga paariviisiline võrdlus kinnitab asjaolu, et ainuke oluline erinevus on esimese laktatsiooni lehmade ja ülejäänud laktatsioonide vahel. Seega on algse nelja taseme asemel laktatsioonid edasises analüüsis grupeeritud kaheks, esimese laktatsiooni loomad ning enam kui esimese laktatsiooni loomad.

Tabel 7. Tunnuseta *Eriliigilised rasvhapped* mudelite parameetrite hinnangud.

Mudel/ Tunnus	30-päevane		45-päevane		60-päevane	
	Hinnang	p-väärtus	Hinnang	p-väärtus	Hinnang	p-väärtus
Vabaliige	0,0176	0,8581	0,0463	0,7232	0,0748	0,6100
<i>De novo</i> rasvhapped	0,1629	0,0001	0,1705	0,0004	0,1596	0,0005
Valmis rasvhapped	-0,0653	0,1369	-0,0989	0,0460	-0,0949	0,0469
Laktatsiooni- päev	-0,0448	0,2559	-0,0137	0,7569	-0,0139	0,7473
Laktatsioon 2	-0,3092	0,0008	-0,4611	<0,0001	-0,4051	0,0001
Laktatsioon 3	-0,2807	0,0138	-0,3292	0,0126	-0,4078	0,0017
Laktatsioon >3	-0,3324	0,0169	-0,4656	0,0052	-0,4159	0,0147

Erineva ajavahemikuga kontroll-lüpside andmetel hinnatud mudeleid omavahel võrreldes ei saa enam Akaike informatsioonile toetuda, sest mudelite hindamisel kasutatav informatsioonikogus on erinev. AIC asemel on mudelite võrdluseks kasutatud graafilist analüüsimeetodit, ROC-kõveraid ja nende aluseid pindalasid. Jooniselt 7 on näha, et mudeli prognoosimisvõime teeb 30-päevase intervalli laiendamisel märkimisväärse hüppe, kõveraalne pind-

ala tõuseb kuni 0,15 ühiku võrra ning on 45- ja 60-päevaste andmestike puhul ligilähedane täiuslikule tulemusele ehk väärtusele 1.



Joonis 7. Tunnuseta *Eriliigilised rasvhapped* peamõjudega mudelite ROC-kõverad.

Kõveraaluse pindala põhjal on parima ennustusvõimega 45 päeva andmetel hinnatud mudel. Selle mudeli parameetrite hinnanguid uurides näeme, et olulisuse nivool 0,05 on statistiliselt olulised tunnused *De novo rasvhapped*, *Valmis rasvhapped* ning tunnuse *Laktatsioon* teise taseme erinevus baastasemest. Valitud olulisuse nivool osutus ainsana statistiliselt ebaoluliseks laktatsiooni päeva mõju tiinestumisele. (Tabel 8)

Parameetrite väärtuste tõlgendamisel tuleb arvestada standardiseerimisega (v.a. tunnuse *Laktatsioon* korral) ehk muutused vastavad ühe standardhälbe suurusele kasvule mitte üheühikulisele kasvule. Vaadates hinnatud parameetrite märke, näeme, et ainsana parandab tiinestumise šanssi *de novo* rasvhapete koguse kasv. Ühe standardhälbeline kasv *de novo* rasvhapete koguses suurendab tiinestumise šanssi 1,2 korda. Valmis rasvhapetel on see-eest vas-

tassuunaline mõju, ühe standardhälbe suurune kasv kahandab tiinestumise šanssi 9% võrra.

Tabel 8. 45-päevase mudeli parameetrite hinnangud ja olulisuse tõenäosused.

Argumenttunnus	Parameetri hinnang	p-väärtus
Vabaliige	0,0451	0,7300
<i>De novo</i> rasvhapped	0,1672	0,0005
Valmis rasvhapped	-0,0971	0,0495
Laktatsioonipäev	-0,0128	0,7721
Laktatsioon ≥ 2	-0,4267	<0,0001

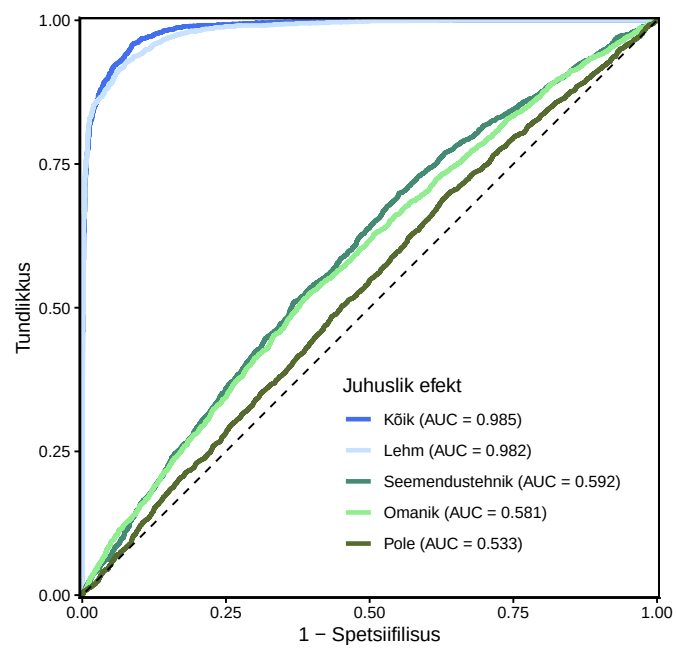
Absoluutväärtuselt suurim mõju on laktatsioonide vahelisel erinevusel. Baas-taseme ehk esimese laktatsiooni loomadel on 1,5 korda suurem šanss tiinestuda kui teise või enama laktatsiooni loomadel. See tulemus on oodatav arvestades asjaolu, et keskmiselt on lehmadel 2,5 laktatsiooni ehk võime tiinestuda langeb peale igat poegimist.

Laktatsioonipäeva mõju tiinestumise šansile on väga marginaalne, tegu on ühtlasi ka ainsa tunnusega, mis ei ole statistiliselt oluline. Ühe standardhälbe suurune kasv laktatsioonipäevas langetab lehma tiinestumise šanssi 1% võrra.

Selleks, et mõista mudeli peaaegu täiuslikku prognoosimisvõimet, tuleb uurida juhuslike efektide mõju prognoosimisele. Joonisel 8 on näha erinevate juhuslike efektidega mudelite selget grupeerumist. Kõige parem kõveraalne pindala, 0,985, on terviklikul mudelil, kuhu on kaasatud kõik kolm juhusliku efekti. Ainult 0,003 ühiku võrra madalam kõveraalne pindala on mudelil, kus ainsa juhusliku efektina on lehma mõju. Nende kahe mudeli võrdlemisi sarnane prognoosimisvõime viitab sellele, et terviklik mudel kasutab prognoosimisel peamise infona iga üksiku lehma unikaalsust. Selle probleemi aluseks võib olla asjaolu, et mudeli hindamiseks kasutatavas andmestikus on iga unikaalse lehma kohta liiga vähe vaatlusi, 3349 lehma ja 5309 vaatlust. Seejuures

on 56,7% lehmadel mõõdetud ainult ühe kontroll-lüpsi tulemus. Ülejäänud kahe juhusliku efekti korral on näha märkimisväärselt langust prognoosimisvõimes, kõveraalsed pindalad on 0,6 ligidal. Seemendustehniku ja omaniku juhusliku efekti minimaalne panus ennustusvõimesse kinnitab veelgi enam, et terviklik mudel õpib unikaalseid lehmi eristama. Kusjuures juhuslike efektidega mudel on prognoosimiseks praktiliselt kasutu, olles kõveraalse pindalaga 0,533 napilt parem täiesti juhuslikust prognoosist. Agnieszka Otwinowska-Mindur jt (Otwinowska-Mindur *et al.*, 2022) hindasid oma uuringus, et poola holsteini tõugu lehma tiinestumise võime geneetiline päritavus jääb sõltuvalt laktatsioonijärgust vahemikku 2,1–2,9%. Sarnase tulemuseni jõudsid ka Wondimu A. Lombebo jt (Lombebo *et al.*, 2026), kes hindasid hiina holsteini lehmadel tiinestumisvõime geneetiliseks päritavuseks ligikaudu 4,4%. Mudelis olev lehma juhuslik efekt arvestabki just iga üksiku lehma geneetilise eripäraga tiinestumise hindamisel. Niisiis pole hinnatud mudelis lehma juhusliku efekti nii suur mõju prognoosimisvõimele kooskõlas Otwinowska-Minduri jt ning Lombebo jt uuringutega ning tegu peab olema andmestruktuuri eripäraga.

Sellegipoolest näitab *de novo* ja valmis rasvhapete statistiline olulisus, et neil rasvhappegruppidel on mõju tiinestumisele. Arvestades aga ka juhuslike efektide mõjudega, tuleb nentida, et pelgalt rasvhappelise koostise teadmine ei võimalda tiinestumist täpselt prognoosida.



Joonis 8. ROC-kõverad erinevate juhuslike efektide korral.

Kokkuvõte

Bakalaureusetöö eesmärk oli uurida kas ja kuivõrd on võimalik kontroll-lüpsil määratud rasvhapete tulemuste põhjal ennustada lehma tiinestumist esimesel seemendusel.

Analüüsis kasutati juhuslike faktoritega logistilist regressioon, kus uuritav tunnus oli lehma tiinestumine esimesel seemendusel (jah/ei). Mudeli fikseeritud argumenttunnused olid *de novo*, eriliigiliste ja valmis rasvhapete kogused, laktatsioonipäev ning laktatsioonijärk. Juhuslike efektidena kaasati mudelisse lehma unikaalsus, seemendustehnik ja omanik ehk farm.

Mudelid hinnati kolmel erineval andmestikul, ROC-kõvera aluse pindala põhjal osutus neist parimaks mudel, kus kontroll-lüpsid toimusid kuni 45 päeva enne esimest seemendust. Eelneva analüüsi tulemusena jäi lõplikust mudelist kõrvale eriliigiliste rasvhapete tunnus. Kasutatud argumenttunnustest osutus ainsana statistiliselt ebaoluliseks laktatsioonipäev. *De novo* ja valmis rasvhapete ning laktatsioonijärgu mõjud olid statistiliselt olulised, kusjuures absoluutväärtuselt suurim mõju oli laktatsioonijärgul.

Tulemuste analüüsimisel ilmnas aga, et mudeli prognoosimisvõime tuleb peaaegu täielikult üksiku lehma unikaalsusest. Ilma lehma juhusliku efektita oli mudeli prognoosimisvõime kasin, kõveraallane pindala ligikaudu 0,6. Juhuslike efektideta mudel oli aga prognoosimiseks üldse kasutu, olles napilt parem juhuslikust prognoosist.

Edasiarendusena võiks kvaliteetsemate andmete peal uurida kitsamas ajaaknas (nt kuni 7 päeva enne seemendust) mõõdetud kontroll-lüpsi tulemuste mõju tiinestumisele. Seejuures võib kasuks tulla rasvhapete absoluutse koguse asemel nende osakaalude kasutamine. Prognoosimise eesmärgil võiks hinnata ja võrrelda ka masinõppe algoritme (nt juhumetsad).

Kasutatud allikad

- Agresti, A. (2002). *Categorical Data Analysis*. 2. väljaanne. John Wiley Sons, lk. 165–183, 491–502.
- Carter, J. V., J. Pan, S. N. Rai ja S. Galandiuk (2016). „ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves“. *Surgery* 159 (6), lk. 1638–1645. DOI: [10.1016/j.surg.2015.12.029](https://doi.org/10.1016/j.surg.2015.12.029).
- Churakov, M., J. Karlsson, A. Edvardsson Rasmussen ja K. Holtenius (2021). „Milk fatty acids as indicators of negative energy balance of dairy cows in early lactation“. *Animal* 15 (7), lk. 1–10. DOI: [10.1016/j.animal.2021.100253](https://doi.org/10.1016/j.animal.2021.100253).
- De Marchi, M., V. Toffanin, M. Cassandro ja M. Penasa (2014). „Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits“. *Journal of Dairy Science* 97 (3), lk. 1171–1186. DOI: [10.3168/jds.2013-6799](https://doi.org/10.3168/jds.2013-6799).
- Gomez, R. S., A. R. Sanchez, C. G. Garcia ja J. G. Perez (2020). „The VIF and MSE in Raise Regression“. *Mathematics* 8 (4), lk. 605. DOI: [10.3390/math8040605](https://doi.org/10.3390/math8040605).
- Hosmer Jr., D. W., S. Lemeshow ja R. X. Sturdivant (2013). *Applied Logistic Regression*. 3. väljaanne. John Wiley Sons, lk. 17–19, 96.
- Lombebo, W. A., H. Kang, D. Mingxin, Z. T. Wondie, G. M. Tarekegn ja H. Zheng (2026). „Genetic parameter estimates and genetic trends for reproductive traits of Holstein dairy cattle in China“. *Journal of Dairy Science* 109 (2), lk. 1775–1783. DOI: [10.3168/jds.2025-27409](https://doi.org/10.3168/jds.2025-27409).
- McCulloch, C. E. ja S. R. Searle (2001). *Generalized, Linear, and Mixed Models*. John Wiley Sons, lk. 220–222.
- McGuire, M. A. ja D. E. Bauman (2002). „Milk Biosynthesis and Secretion“. Teoses: *Encyclopedia of Dairy Sciences*. Toim. H. Roginski, J. W. Fuquay ja P. F. Fox. Köide 3. Elsevier Science Ltd, lk. 1828–1834.

- Otwinowska-Mindur, A., E. Ptak, W. Jagusiak ja A. Zarnecki (2022). „Estimation of Genetic Parameters for Female Fertility Traits in the Polish Holstein-Friesian Population“. *Animals* 12 (12), lk. 1485. DOI: [10.3390/ani12121485](https://doi.org/10.3390/ani12121485).
- Pentjärv, A ja M. Henno (2025). *Piima rasvhappeline profiil - abivahend lüpsilehmade söötmise optimeerimisel*. Eesti Põllumajandusloomade Jõudluskontroll. URL: <https://www.epj.ee/jkk/piimaveised/piimaveiste-j%C3%B5udluskontrolli-kasulik-teave/epj-s%C3%B5numid-74/piima-rasvhappeline-profiil.html> (vaadatud 12.11.2025).
- Schwarz, D., M. R. Bak ja P. W. Hansen (2022). „Development of global fatty acid models and possible applications“. *International Journal of Dairy Technology* 75 (1), lk. 4–20. DOI: [10.1111/1471-0307.12820](https://doi.org/10.1111/1471-0307.12820).
- Wang, X., V. Liesaputra, Z. Liu, Y. Wang ja Z. Huang (2023). „An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification“. *Artificial Intelligence In Medicine* 147. DOI: [10.1016/j.artmed.2023.102738](https://doi.org/10.1016/j.artmed.2023.102738).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Markus Erik Tommula,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Kas ja kuivõrd ennustab piima rasvhappeline koostis lehma tiinestumist?", mille juhendaja on Tanel Kaart, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Markus Erik Tommula

12.05.2026