

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Karoliine Inselberg
**Eesti holsteini tõugu piimalehmade
tiinestumine hinnatuna nende piima erinevat
tüüpi soomaatiliste rakkude suhtarvu ja
koguarvu alusel**

Matemaatiline statistika
Bakalaureusetöö (9 EAP)

Juhendaja: PhD Tanel Kaart

TARTU 2025

**EESTI HOLSTEINI TÕUGU PIIMALEHMADE TIINESTUMINE
HINNATUNA NENDE PIIMA ERINEVAT TÜÜPI SOMAATILISTE
RAKKUDE SUHTARVU JA KOGUARVU ALUSEL**

Bakalaureusetöö

Karoliine Inselberg

Lühikokkuvõte. Bakalaureusetöö eesmärgiks on uurida, kas ja kuidas mõjutab seemenduse eelselt toimunud kontroll-lüpsil määratud somaatiliste rakkude arv ja erinevat tüüpi somaatiliste rakkude suhtarv lehma tiinestumist esimesel seemendusel. Töö koosneb kolmest peatükist, millest esimeses on antud ülevaade somaatiliste rakkude olemusest, teises peatükis on kirjeldatud analüüsis kasutatavat mudelit ja kolmandas peatükis on esitatud Eesti piimalehmade andmete analüüsi tulemused..

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

Märksõnad: logistiline mudel, piimalehmad, tiinestumine, udara tervis.

**CONCEPTION OF ESTONIAN HOLSTEIN DAIRY COWS
EVALUATED BASED ON THE RATIO AND TOTAL COUNT OF
DIFFERENT TYPES OF SOMATIC CELLS IN MILK**

Bachelor thesis

Karoliine Inselberg

Abstract The aim of this bachelor's thesis is to investigate whether and to what extent the somatic cell count and the ratio of different types of somatic cells, determined during the pre-insemination test milking, affect a cow's conception at the first insemination. The thesis consists of three chapters,

the first of which provides an overview of the nature of somatic cells, the second chapter describes the model used in the analysis, and the third chapter presents the results of Estonian dairy cows' data analysis.

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

Key Words: logistic model, dairy cows, reproduction, udder health.

Sisukord

Sissejuhatus	5
1 Somaatilised rakud lehmas piimas	6
1.1 Somaatiliste rakkude arv	6
1.2 Somaatiliste rakkude eristamine	7
1.3 Udara tervise klassid	7
2 Logistilise regressioonanalüüsi mudel	9
2.1 Fikseeritud faktoritega logistiline mudel	9
2.2 Juhuslike faktoritega logistiline mudel (üldistatud segamudel) . . .	12
2.3 ROC-kõver	13
3 Eesti piimalehmade andmete analüüs	15
3.1 Andmed ja nende ettevalmistus analüüsiks	15
3.2 Andmete esmane kirjeldav analüüs	16
3.3 Lehmade tiinestuvuse hindamine udara tervise gruppide alusel . . .	26
Kokkuvõte	33
Lisa 1	34
Lisa 2	35
Lisa 3	36
Kasutatud allikad	37

Sissejuhatus

Selleks, et lehm piima annaks, peab ta poegima. Poegimise eeldus omakorda on tiineks jäämine. Eestis kulub ühe lehma tiinestamiseks keskmiselt 1,9 seemendust (EPJ, 2024) ning lehmade mittetiinestumine on üks suuremaid karja taastootmise ja piimafarmide majandusliku toimetuleku murekohti. Üheks lehmade mittetiinestumise põhjuseks on poegimisejärgselt (sageli seoses suure piimatoodanguga) tekkinud terviseprobleemid. Piima somaatiliste rakkude arvu on aastakümneid kasutatud piima kvaliteedi ja lehmade udarapõletiku markerina. Uuem laboritehnoloogia võimaldab lisaks somaatiliste rakkude üldarvule leida ka erinevat tüüpi somaatiliste rakkude osakaalu piimaproovides ning Eesti Põllumajandusloomade Jõudluskontrolli AS-i piimalaboris on sellisel viisil analüüsitud kõigi lehmade viimase kahe aasta igakuised kontroll-lüpsid.

Käesoleva bakalaureusetöö eesmärgiks on uurida, kas ja kuivõrd on võimalik seemendamise eelselt toimunud kontroll-lüpsil määratud somaatiliste rakkude koguarvu ja erinevat tüüpi somaatiliste rakkude suhtarvu alusel ennustada lehmade tiinestumist esimesel seemendusel.

Bakalaureusetöö koosneb kolmest suuremast peatükist. Esimeses peatükis on antud ülevaade somaatilistest rakkudest lehma piimas - mis need on, kuidas neid eristatakse ja kuidas jagunevad somaatiliste rakkude arvu ja somaatiliste rakkude suhtarvu alusel lehmad udara tervise gruppidesse. Teises peatükis on kirjeldatud analüüsis kasutusel olnud logistilist regressioonmudelit. Ülevaade on tehtud nii fikseeritud faktoritega logistilisest mudelist kui ka juhuslike faktoritega logistilisest mudelist. Lühidalt on juttu ka ROC-kõverast. Kolmandas peatükis on läbi viidud eesti piimalehmade andmete analüüs.

1 Somaatilised rakud lehma piimas

Somaatilised rakud on segu valgetest verelibledest ehk leukotsüütidest ning piimäärme epiteelirakkudest. Valged verelibled moodustavad somaatiliste rakkude arvust 98-99% ja nende ülesandeks piimas on udarapõletikku põhjustavate mikroobide äratundmine ja nende hävitamine. (ICBF, 2012) Leukotsüütide hulka kuuluvad mitmed immuunsüsteemi rakud: lümfotsüüdid, makrofaagid ja polümorftuumalised neutrofiilid. Neil kõigil on udaras erinev ülesanne, aga mikroobide hävitamisega tegelevad peamiselt polümorftuumalised neutrofiilid. (Lillik *et al.*, 2021)

Kui haigustekitajad on nisa kaudu udarasse pääsenud, reageerib organism koheselt udarapõletikuga, millest üks levinumaid avaldumise viise on mastiidi kujul. (Kalmus, 2019) Leukotsüütide arv udaras pole kunagi null, sest makrofaagide ja lümfotsüütide ülesandeks on ka pidevalt potentsiaalselt haigustekitavaid mikroobe otsida. Kui udaras tuvastatakse mõni bakter, saavad polümorftuumalised neutrofiilid koheselt signaali nende hävitamiseks, mistõttu nende arv ja sestap ka somaatiliste rakkude arv kasvab. (Lillik *et al.*, 2021)

1.1 Somaatiliste rakkude arv

Somaatiliste rakkude arvu (SRA) all mõistetakse somaatiliste rakkude arvulist kogust ühes milliliitris piimas.

Lehma udar koosneb miljonitest piima tootvate rakkude rühmadest, mida ühiselt nimetatakse alveoolideks. Alveoolis asuvas luumenis koguneb piim, mida saab piimajuhade kaudu transportida ja hiljem nisade kaudu väljutada. Kui udaras on infektsioon, kogunevad allveoolidesse polümorftuumalsed neutrofiilid. Selleks, et pääseda ligi bakterite kasuvukohta, tuleb neil tungida ühenduskohtadesse, kus asuvad piima tootvad rakud, lükates need rakud oma kohalt välja. Selline protsess toob endaga kaasa vähenenud piima tootmise. (ICBF, 2012) Kui lehm on terve ja infektsioonid puuduvad, on somaatilisi rakke udaras vähe. Kuigi lehma piima SRA sõltub

mitmetest erinevates faktoritest, on siiski leitud, et tervel lehmal on udaras soomaatiliste rakkude arv valdavalt alla 100 000 raku ml-s. Mitmete teadusuuringutega on tõestatud, et kui SRA on suurem kui 200 000 rakku ml-s, on 75%-l juhtudest lehmal vähemalt üks udaraveerand nakatunud, ja kui SRA ületab ühes milliliitris 150 000 raku piiri, on lehmal vähemalt üks udaraveerand nakatunud 65%-l juhtudest. Kuna suurenenud SRA mõjutab peale udara tervise ka muid näitajaid (näiteks piimatoodang), siis kasutab Eesti Põllumajandusloomade Jõudluskontrolli AS (EPJ) 150 000 rakku/ml piirarvuna, et võimalikku põletikku varem tuvastada ja selle abil karja udaratervist parandada. (Kalmus, 2019)

1.2 Somaatiliste rakkude eristamine

Kuigi suurenenud soomaatiliste rakkude arv on olnud üheks peamiseks udarapõletiku indikaatoriks juba aastaid, ei saa üldarvu kaudu eristada piimas esinevaid soomaatilisi rakke, millel on põletiku arenemisel erinev roll. Nakkusvabas udaras on SRA väike ning peamiseks soomaatilisteks rakkudeks on makrofaagid ja lümfotsüüdid. Kui udar aga nakatub, suureneb ning saavutab märkimisväärse ülekaalu polümorftuumaliste neutrofiilide (PMN) arv. (EPJ, i.a)

Somaatiliste rakkude eristamine (SRE) näitab seda, kui suure osa kõigist leukotsüütidest moodustavad polümorftuumalised neutrofiilid ja lümfotsüüdid. Kui PMN ülesandeks on mikroobide hävitamine ja udarakoe taastamine, siis lümfotsüütide peamiseks ülesandeks on antikehade tootmine. (EPJ, i.a)

1.3 Udara tervise klassid

Kui kasutada nii SRA-d kui ka SRE-d, saab neid kombineerides jagada lehmad piimaproovi võtmise päeva (so kontrollpäeva) seisuga nelja udaratervise gruppi (UTG) (Lillik *et al.*, 2021):

A - terved loomad: $SRA \leq 150$ ja $SRE \leq 65$;

B - haiguseelne seisund/oht haigestuda: $SRA \leq 150$ ja $SRE > 65$;

C - tõenäoliselt nakatunud lehmad (põletiku aktiivne faas): $SRA > 150$ ja $SRE > 65$;

D - tõenäoliselt krooniline udarapõletik: $SRA > 150$ ja $SRE \leq 65$.

Seejuures järjestatakse UTG-d parem-halvem skaalal: $A > B > C > D$.

Eesti Põllumajandusloomade Jõudluskontrolli ASi infolehes kirjutatu alusel võiks hea udaratervisega karjas olla esimeses ehk A-grupis vähemalt 70% lehmadest, B-grupis vähem kui 20%, C-grupis vähem kui 10% ja D-grupis alla 2% lehmadest. (Lillik *et al.*, 2021) Kõige olulisemad ja enim informatsiooni andvad grupid on B ja D, sest nendes saab SRE alusel tuvastada loomad, kel on oht haigestuda või kelle udarapõletik on suure tõenäosusega krooniline. Kui somaatilisi rakke ei eristataks, poleks nende faaside tuvastamine võimalik. (Lillik *et al.*, 2021)

Varasemalt on uuritud kirjeldatud viisil moodustatud UTG-de seoseid mitmete näitajatega. Näiteks selgus ühes teadustöös, et UTG-s A oli piimatoodang kõige kõrgem ja grupiti edasi liikudes piimatoodang vähenes märkimisväärselt. Eesti andmete kohaselt oli piimatoodang võrreldes grupiga A grupis B 2,4% väiksem, grupis C 9,8% väiksem ja grupis D juba 38,5% väiksem. Lisaks piimatoodangule täheldati udara tervisel seost ka piima kuivainega, hooajaga ja lüpsisagedusega. (Schwarz *et al.*, 2020)

2 Logistilise regressioonanalüüsi mudel

Logistiline mudel ehk logistiline regressioon on statistiline meetod, mida kasutatakse binaarse (jah/ei) sõltuva muutuja modelleerimiseks. Logistilise regressiooni eesmärk on hinnata tõenäosust, et mingi sündmus toimub, tuginedes sõltumatu(te)le muutuja(te)le ehk faktori(te)le. Antud peatükis on mudeli kirjeldamiseks kasutatud Alan Agresti raamatut „Categorical Data Analysis” (Agresti, 2002), Liina-Mai Toodingu raamatut „Andmete analüüs ja tõlgendamine sotsiaalteadustes” (Tooding, 2007) ja Charles E. McCullochi ning Shayle R. Searle'i raamatut „Generalized, Linear and Mixed Models” (McCulloch ja Searle, 2001).

2.1 Fikseeritud faktoritega logistiline mudel

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

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

Siis lihtne logistiline mudel avaldub kujul

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

kus α on vabaliige, x argumenttunnuse X väärtus ning β argumenttunnuse mõju tugevust ja suunda väljendav regressioonikordaja. Regressioonikordaja ees olev märk näitab seda, kas sündmuse toimumise tõenäosus x -i kasvades suureneb või väheneb. Taolise funktsiooni väärtused jäävad alati nulli ja ühe vahele.

Lineaarse, kogu arvteljel pideva funktsioonina on logistiline mudel esitatav kujul

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

Selles esituses $\pi(x)$ on sündmuse toimumise tõenäosus ja $1 - \pi(x)$ on sündmuse mittetoimumise tõenäosus ning suhe

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

on defineeritud kui sündmuse toimumise šanss. Šansi tarvis on logistiline mudel esitatav kujul

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

ning sellest esitusest 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, kui mitu korda muutub sündmuse toimumise šanss argumenttunnuse X väärtuse suurenemisel ühe võrra:

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

millest

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

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

Logistiline regressioon on laiendatav ka mitme argumenttunnusega juhule. Olgu $\pi(x) = P(Y = 1)$ sündmuse toimumise tõenäosus väärtustel $x = (x_1, \dots, x_p)$ p muutujaga. Sel juhul avaldub logistiline mudel järgnevalt:

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

ja tõenäosus $\pi(x)$ avaldub kujul

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

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 , on mudel esitatav sarnaselt mitmese logistilise regressiooni mudelile kujul

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

kus

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

on tunnuse X i -nda taseme x_i indikaatorfunktsioon ja β_i on selle taseme mõju.

Kuna diskreetse faktortunnuse puhul võetakse parameetrite ühese hinnangu saamiseks enamasti üks tase (näieks x_1) baasiks ehk referentstasemeks ja sellele vastav mõju võrdseks nulliga ($\beta_1 = 0$), iseloomustab mudeli vabaliige sündmuse toimumise šanssi baastasemel:

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

ja ülejäänud tasemete mõjud näitavad erinevust baastasemest. Seejuures kujutab suhe

$$\frac{\exp[\alpha + \beta_i \mathbb{I}(X = x_i)]}{\exp[\alpha + \beta_1 \mathbb{I}(X = x_1)]} \Big|_{\beta_1=0} = \frac{\exp(\alpha) \cdot \exp(\beta_i)}{\exp(\alpha)} = \exp(\beta_i)$$

enesest šansside suhet

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

ja seega 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 (üldistatud segamudel)

Erinevalt fikseeritud faktoritest, mille tasemed esindavad konkreetseid eelnevalt valitud ja/või uuringut korrates taasvalitavaid/fikseeritavaid omadusi, on juhuslike faktorite potentsiaalsete tasemete arv väga suur (lõpmatu). Andmetes on neist esindatud juhuslik valik ning uuringut korrates ei ole võimalik kaasata kõigi samade tasemete mõjusid.

Juhusliku faktori tasemete mõjud eeldatakse tulenevat suure hulga otseselt mitte-mõõdetavate väikeste mõjutegurite summast, mistõttu eeldatakse modelleerimisel, et juhusliku faktori tasemete mõjud jaotuvad normaaljaotuse järgi keskmisega 0 ja mingi dispersiooniga, mis iseloomustab juhusliku faktori mõju suurust.

Olgu X diskreetne fikseeritud faktortunnus tasemetega x_1, \dots, x_k ja tasemete mõjudega β_1, \dots, β_k ning Z diskreetne juhuslik faktortunnus andmetes esindatud tasemetega z_1, \dots, z_t ja tasemete mõjudega u_1, \dots, u_t , kus $u_j \sim N(0, \sigma_z^2)$, $j = 1, \dots, t$. Kuna juhusliku faktori tasemete keskmine mõju on null, avaldub uuritava sündmuse tõenäosus kui uuritava tunnuse y keskvärtus vaid fikseeritud faktori tasemete mõjude kaudu:

$$\pi(x, z) = \pi(x) = \frac{\exp[\alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k)]}{1 + \exp[\alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k)]}.$$

Kui võtta sündmuse tõenäosuse hindamisel arvesse ka juhusliku faktori realiseerunud mõjud, on tulemuseks uuritava sündmuse tinglik tõenäosus:

$$\pi(x|z) = \frac{\exp[\alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k) + u_1 \mathbb{I}(Z = z_1) + \dots + u_t \mathbb{I}(Z = z_t)]}{1 + \exp[\alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k) + u_1 \mathbb{I}(Z = z_1) + \dots + u_t \mathbb{I}(Z = z_t)]}$$

ehk logistiline mudel kujul

$$\text{logit}[\pi(x|z)] = \alpha + \beta_1 \mathbb{I}(X = x_1) + \dots + \beta_k \mathbb{I}(X = x_k) + u_1 \mathbb{I}(Z = z_1) + \dots + u_t \mathbb{I}(Z = z_t).$$

2.3 ROC-kõver

Antud alapeatüki kirjeldamiseks on kasutatud Francis Sahngun Nahm'i teadusartiklit „Receiver operating characteristic curve: overview and practical use for clinicians” (Nahm, 2022) ja Tanel Kaarti koostatud õpiobjekti „Binaarsete tunnuste analüüsimeetodid” (Kaart, 2012).

Kasutades mingi sündmuse prognoosimiseks mudeleid, on üheks olulisemaks kriteeriumiks mudeli täpsus. Binaarse tunnuse korral saab prognoosi tulemuse jaotada nelja erineva kategooria vahel:

- tõeselt negatiivne (TN),
- valenegatiivne (FN),
- tõeselt positiivne (TP),
- valepositiivne (FP).

Õigete või valede positiivsete ning negatiivsete tulemuste abil on võimalik leida palju erinevaid näitajaid prognoosimaks mudeli täpsust. Nendest ühed enam kasutatavad näitajad on tundlikkus ja spetsiifilisus.

Tundlikkus kirjeldab, kui suur osa tegelikult toimunud sündmustest tuvastatakse õigesti ehk teisisõnu, kui hästi ennustab mudel sündmuse esinemist. Tundlikkuse saab leida suhtelise sageduse abil, jagades tõeselt positiivsete tulemuste arvu tõeselt positiivsete ja valenegatiivsete tulemuste summaga:

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

Spetsiifilisus näitab aga seda, kui suur osa sündmuse mittetoimumisest on õigesti tuvastatud ehk kui hästi suudab mudel ära tunda mittetoimunud juhtumeid.

Spetsiifilisuse leidmiseks tuleb tõeselt negatiivsete juhtude arv jagada tõeselt negatiivsete ja valepositiivsete juhtude sumaga:

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

ROC-kõveraks nimetatakse graafiliselt kujutatud analüütilist meetodit, mille abil hinnatakse prognoosi täpsust erinevatele argumentide väärtustele vastavate tundlikkuse ja spetsiifilisuse paaride esitusel. Tundlikkust on joonisel enamasti kujutatud y-teljel ning x-teljel kujutatakse 1-spetsiifilisuse ehk valepositiivsete määra väärtused.

ROC-kõvera abil saab leida ROC-kõvera aluse pindala (AUC), mis on üheks mudeli headuse näitajaks. Juhul kui sündmuse ennustuse täpsus on juhuslik ehk 50%, vastab AUC väärtusele 0,5, mis tähendab seda, et mudeli argument prognoositava sündmuse toimumisel mingit rolli ei mängi. Mida rohkem erineb AUC 0,5-st, seda täpsem antud mudel oma ennustustes on. Kokkuleppeliselt on mudeli headust näitavad piirid järgmised:

- kui $0,5 \leq AUC < 0,6$, pole mudeli abil mõtet üldse midagi prognoosida,
- kui $0,6 \leq AUC < 0,7$, on mudeli täpsus kasin,
- kui $0,7 \leq AUC < 0,8$, on mudeli täpsus rahuldav,
- kui $0,8 \leq AUC < 0,9$, on täpsus hea,
- kui $0,9 \leq AUC$, on täpsus suurepärase.

3 Eesti piimalehmade andmete analüüs

3.1 Andmed ja nende ettevalmistus analüüsiks

Uuringus kasutatud andmed on pärit Eesti Põllumajandusloomade Jõudluskontrolli ASi andmebaasist. Analüüsiks on kasutatud rakendustarkvara R.

Kasutatud on nende eesti holsteini tõugu piimalehmade andmeid, kes on poeginud ajavahemikus 01.06.2022 kuni 01.06.2024. Selles vahemikus on toimunud kokku 157 949 poegimist.

Sõltuvaks tunnuseks on lehma tiinestumine esimese seemenduse tulemusel: jah/ei (1/0) ja oluliseimaks argumenttunnuseks on esimese seemenduse eelselt toimunud kontroll-lüpsi SRA ja SRE alusel määratud UTG. Kuna kontroll-lüpsid toimuvad kord kuus ja esimene seemendus ei toimu tavaliselt varem kui 60 päeva peale poegimist, on paljudel lehmadel seemendamise eelne UTG määratud korduvalt. Seetõttu on antud töös tiinestumist seostatud kolmel eri viisil määratud UTG-ga. Nendest esimese puhul on UTG määratud esimese poegimisjärgse kontroll-lüpsi SRA ja SRE põhjal. Teisel juhul on UTG määratud esimesele seemendusele eelnenud lähima kontroll-lüpsi SRA ja SRE alusel. Ja kolmandal juhul on võetud esimesele seemendusele eelnenud kahe kontroll-lüpsi kehvem UTG.

Järgnevalt on toodud välja lisakriteerimid kõigi kolme variandi kohta.

1. Kui udara tervise grupi määramisel on aluseks poegimisele järgnenud esimene kontroll-lüps.
 - Päevade arv poegimisest esimese kontroll-lüpsini jääb vahemikku 5-45.
2. Kui udara tervise grupi määramisel on aluseks esimesele seemendusele eelnenud lähim kontroll-lüps.
 - Päevade arv poegimisest kontroll-lüpsini peab olema suurem kui 4.

- Päevade arv kontroll-lüpsist esimese seemenduseni jääb vahemikku 1-30.
3. Kui udara tervise grupi määramisel on aluseks esimesele seemendusele eelnenud kahe lähima kontroll-lüpsi kehvem tulemus.
- Päevade arv poegimisest kontroll-lüpsini peab olema suurem kui 4.
 - Päevade arv kontroll-lüpsist esimese seemenduseni jääb vahemikku 1-60 päeva.

Lisaks peab kõikide andmete puhul olema vahemik poegimisest esimese seemenduseni 60-150 päeva.

Ka lehma tiinestumise üle on võimalik otsustada mitmel eri viisil. Antud töös on kasutatud kahte varianti. Esimesel juhul on eelduseks järgnenud poegimise olemasolu, mille abil saab tiinestumise esimesest seemendusest leida olemasolevast andmetabelist. Sellisel juhul märgitakse esimene seemendus tiinestavaks juhul, kui poegimise ja seemenduse vahel on 280 ± 5 päeva. Teisel juhul on vaatluses kõik lehmad, keda pole prakeeritud 45 päeva jooksul peale esimest seemendust ning kui 45 päeva jooksul pole toimunud uut seemendust, loetakse lehm tiineks. Võrreldes esimese variandiga on nüüd andmestikus ka lehmad, kellel pole järgmist poegimist registreeritud ehk lehm on suure tõenäosusega karjast välja läinud.

Nende tingimuste tulemusel tekib 6 erinevat andmestikku, mille abil tiinestumise sõltuvust udara tervise grupist uuritakse.

3.2 Andmete esmane kirjeldav analüüs

Järgnev kirjeldav analüüs on koostatud andmete põhjal, mille aluseks on esimesele seemendusele eelnenud lähima kontroll-lüpsi abil saadud udara tervise grupid ning kus esimesest seemendusest tiinestumine on määratud selle järgi, kas 45 päeva

jooksul on toimunud uus seemendus või mitte. Ülejäänud viie andmestiku puhul olid tulemused sarnased.

Andmestikus on iga holsteini piimalehma kohta toodud välja tunnus *tiinestumine*, mille väärtus saab olla kas 1 või 0. Kokku on analüüsimiseks saadaval 88 360 seemenduse andmed. Argumenttunnusteid on kokku 4 ja nendeks on:

- udara tervise grupp (UTG) - 4-tasemeline faktortunnus, mis on saadud kontrolllülpsi SRA ja SRE kaudu;
- laktatsioon - 3-tasemeline faktortunnus, mille alusel on lehmad jagatud 1., 2. ja 3. ning enama laktatsiooni lehmadeks;
- päevad - päevade arv poegimisest esimese seemenduseni ehk taastumisperioodi pikkus, tunnus on jaotatud kaheks grupiks: 60 – 90 päeva ning 91 – 150 päeva;
- suguselekteeritud sperma kasutamine - jah/ei tunnus.

Tabelis 1 on kujutatud udara tervise gruppidesse jaotumine. Nagu näha, on ligi 70% lehmadest terved loomad ehk asetsevad grupis A. Gruppidesse B ja C on jaotatud üsna võrdselt, kus osakaalud on vastavalt 15,3% ja 14,3% ning oluliselt vähem on kroonilisi haigeid (grupp D), kus osakaal on vaid 0,5%. Olgugi et heaks näitajaks peetakse seda, kui grupis C on alla 10% lehmadest, on üldine jaotus üsna tervislik ja hea.

Tabel 1: Udara tervise gruppidesse jaotumine.

Udara tervise grupp	A	B	C	D	Kokku
Sagedus	61731	13538	12611	480	88360
Protsentuaalne jaotus	69,9	15,3	14,3	0,5	100,0

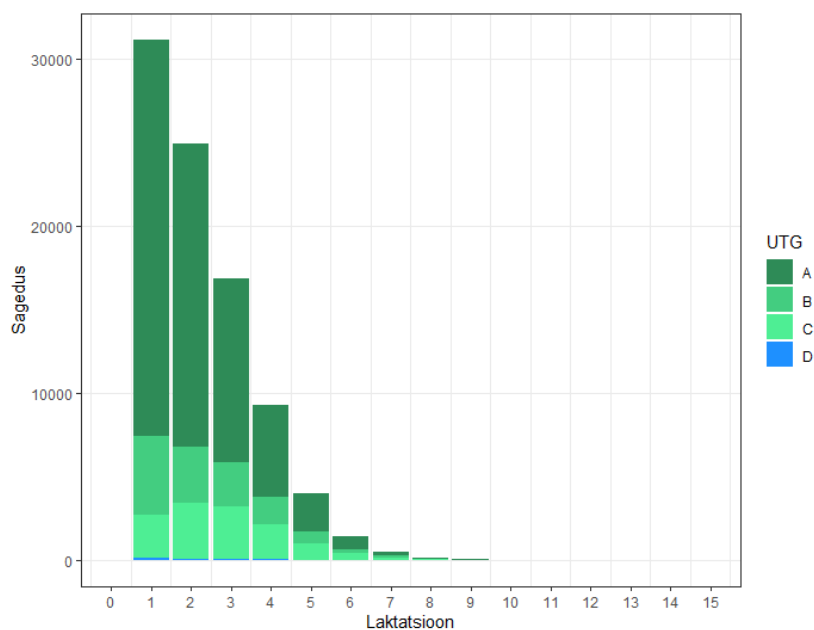
Järgnevalt on vaadatud udara tervise gruppide jaotumist argumenttunnusena kasutatavate faktorite eri tasemetel. Vaadates lehmade vanuselist jaotust, on kõige

rohkem 3. ja vanema laktatsiooni piimalehmi (36,6%). 1. laktatsiooni lehmi on ligi ühe protsendipunkti võrra vähem ning 2. laktatsiooni lehmi kõige vähem, jäädes 28% lähistesse. Protsentuaalselt on vanemaid lehmi rohkem halvemates udara tervise gruppides. Tabelis 2 on näha, kuidas grupis C olevate lehmade jaotus kasvab märgatavalt vanuse kasvades. 3. ja vanema laktatsiooni lehmade hulgas on võrreldes noorematega lehmi rohkem ka udara tervise gruppides B ja D. Suurim erinevus ongi 1. laktatsiooni A grupi lehmade ja 3. ning vanema laktatsiooni A grupi lehmade jaotuse vahel, kus lehma vanemaks saamisel terveid loomi on ligi 15 protsendipunkti võrra vähem. Ja vastupidiselt on jaotumine 3. ja vanemate lehmade hulgas suurem grupis C, kus võrreldes esimese laktatsiooni lehmadega nende osakaal kasvab ligi 12 protsendipunkti.

Tabel 2: Udara tervise gruppidesse jaotumine laktatsioonide lõikes (%) ja laktatsioonide jaotus (seemenduste arv ja osakaal laktatsioonide kaupa).

Udara tervise grupp Laktatsioon	A	B	C	D	Kokku
1	76,12	15,08	8,42	0,38	31 136 (35,24%)
2	72,85	13,44	13,25	0,46	24 924 (28,20%)
3+	61,52	17,01	20,70	0,77	32 300 (36,56%)

Joonisel 1 on näha täpsemalt erinevate laktatsioonide lehmade jaotumist udara tervise gruppidesse. Nagu jooniselt näha, kuuluvad vanusegruppi „3. laktatsioon ja vanemad” peamiselt 3., 4., ja 5. laktatsiooni lehmad ning nende seas on tõepoolest näha grupi C proportsionaalne suurenemine võrreldes grupiga B.



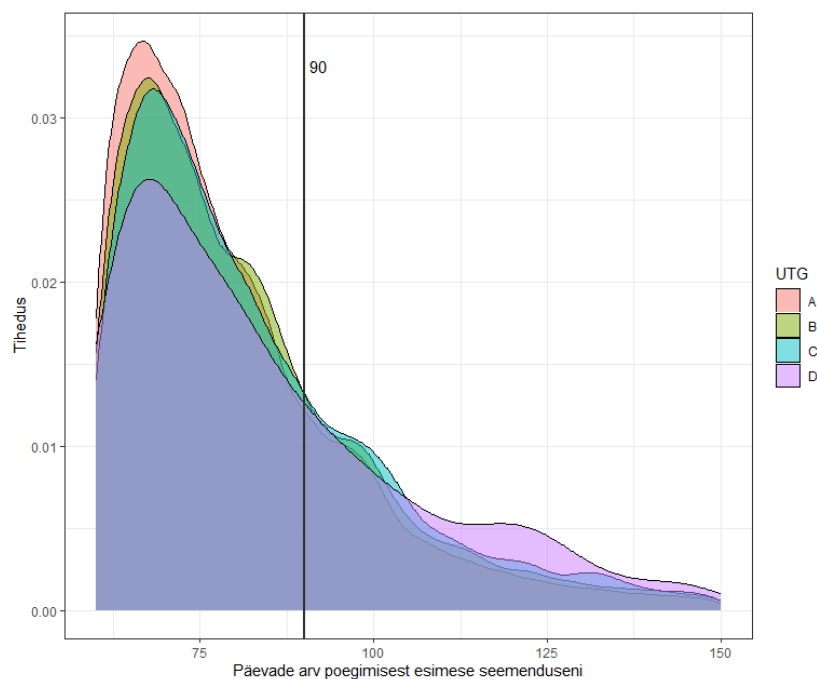
Joonis 1: Udara tervise gruppide jaotus laktatsioonide kaupa.

Vaadates udara tervise gruppidesse jaotumist taastumisperioodi pikkuse lõikes, on selgelt rohkem piimaveiseid rühmas, kus vahemik on 60-90 päeva. Kokku on selles rühmas 77,6% vaatluse all olevatest piimaveistest. Tabelist 3 on näha, et pea 71% lehmadest, kelle taastumisperioodi pikkus jääb vahemikku 60-90 päeva, paigutuvad udara tervise gruppi A. Mingit kindlat seost jaotuste põhjal kahe tunnuse vahel ei ilmne. Igas rühmas olevate lehmade arv pigem kahaneb vastavalt udara tervise grupi halvenemisele, välja arvatud vahemikus 91-150 olevate lehmade arv, kus grupis B ja C on ligikaudu sama suur arv loomi. Siiski on näha väikest langust tervete loomade hulgal taastumisperioodi pikenemisel, kus tervete loomade hulk ligikaudu neli protsendipunkti kahaneb.

Tabel 3: Udara tervise gruppidesse jaotumine sõltuvalt taastumisperioodi pikkusest (%); kokku-veerus on toodud seemenduste arv ja osakaal sõltuvalt taastumisperioodi pikkusest.

Udara tervise grupp/ Taastumisperioodi pikkus	A	B	C	D	Kokku
60-90	70,85	15,07	13,58	0,50	68 534 (77,56%)
91-150	66,44	16,19	16,65	0,72	19 826 (22,44%)

Joonisel 2 on taastumisperioodi pikkus pandud pidevale teljele ning seda udara tervise gruppidesse jaotumisega võrreldud. Nagu jooniselt näha, siis jaotumine on igas grupis väga ühetaoline. Joonisel olev must vertikaalne joon tähistab piiri faktortunnuse kahe taseme vahel, seega kinnitab joonis ka tabelis 3 välja tulnut, et suur osa lehma jääb 60-90 päeva vahemikku. Väikest muutust võib täheldada grupi D osatähtsuse suurenemisel taastumisperioodi pikenemisel.



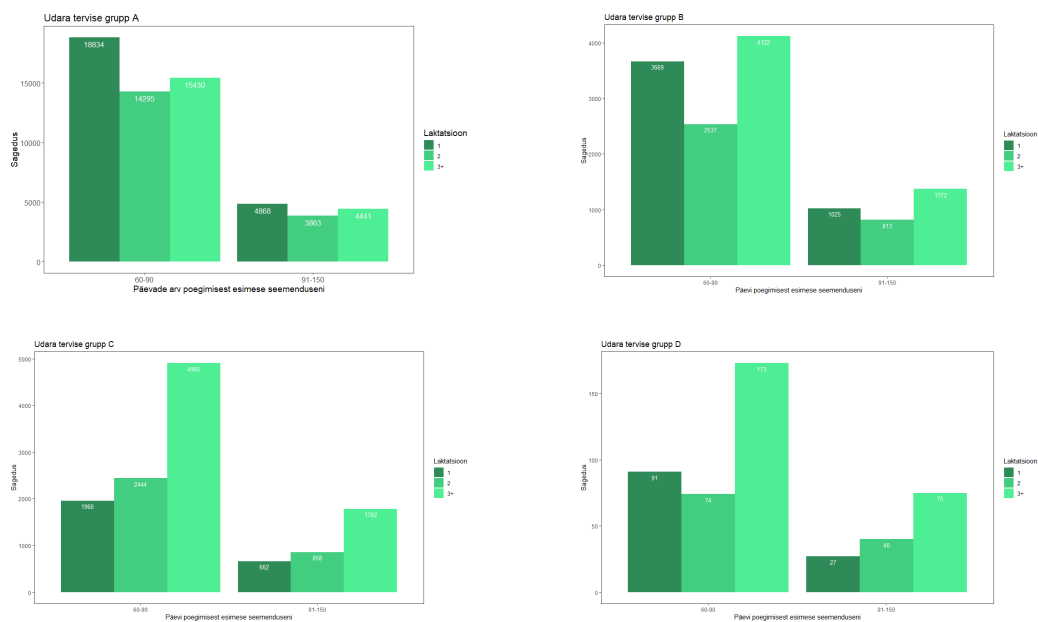
Joonis 2: Poegimisest esimese seemenduseni kulunud päevade arvu jaotus udara tervise gruppide (UTG) kaupa.

Kõikidest seemendamistest kasutakse proportsionaalselt suguselekteeritud spermat üldiselt vähe. See tuleb ka välja antud töö andmete pealt ja tabelis 4 on näha, et suguselekteeritud spermat on kasutatud vaid 6,5% juhtudest. Kuigi suurt erinevust suguselekteeritud sperma kasutamisel ja mittekasutamisel udara tervise gruppidesse jaotumisel ei ole, on näha siiki, et suguselekteeritud spermat kasutatakse rohkem tervetel loomadel.

Tabel 4: Udara tervise gruppidesse jaotumine suguselekteeritud sperma kasutamise lõikes (%); kokku-veerus on toodud seemenduste arv ja osakaal sõltuvalt suguselekteeritud sperma kasutamisest.

Udara tervise grupp/ Suguselekteeritud sperma kasutamine	A	B	C	D	Kokku
ei	69,74	15,33	14,37	0,56	82 602 (93,48%)
jah	71,64	15,18	12,89	0,29	5758 (6,52%)

Võrreldes omavahel laktatsioonide jaotust sõltuvalt taastumisperioodi pikkusest ja UTG-st, on märgata mõningaid erinevusi. Udara tervise grupis A on jaotumine pigem ootuspärane, kus 1. laktatsiooni lehma on kõige rohkem ja 2. laktatsiooni lehma kõige vähem. Kui kõikide gruppide peale kokku on enim küll 3. ja vanema laktatsiooniga lehma, siis võib oletada, et nooremate lehmade parema tervise tõttu neid grupis A rohkem ongi. Grupis B on pilt juba mõnevõrra teine, sest seal on enim lehma just 3. ja vanema laktatsiooni esindajate seas ja seda nii vahemiku 60-90 kui ka 91-150 korral. Võrreldes teistega on 2. laktatsiooni esindajaid oluliselt vähem. Udara tervise grupis C domineerivad kindlalt 3. ja vanema laktatsiooni lehmad, keda on võrreldes teiste laktatsioonidega pea poole võrra rohkem. 1. laktatsiooni lehma on nüüd kokku kõige vähem, mis läheb kooskõlla arvamusega, et nooremad loomad on tervemad. Grupp D on ainus, kus ka taastumisperioodi pikkuse kaupa on märgata erinevust. Kui vahemikku 60-90 jäävate lehmade puhul on jällegi selgelt rohkem 3. ja vanema laktatsiooni lehma ja kõige vähem 2. laktatsiooni lehma, siis vahemikus 91-150 on ikkagi taaskord 1. laktatsiooni lehma kõige vähem ja 3. ning vanema laktatsiooni lehma kõige rohkem.



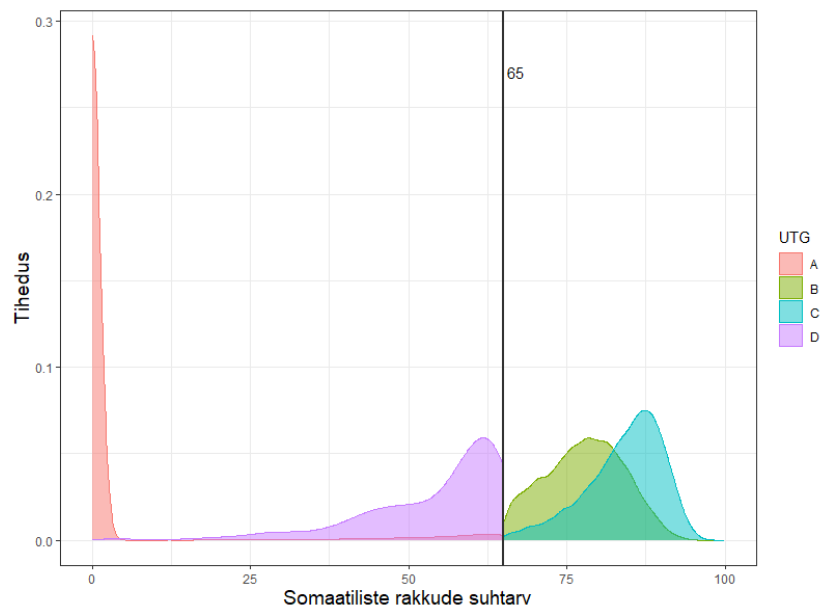
Joonis 3: Laktatsiooni seos taastumisperioodi pikkusega udara tervise gruppide kaupa.

Uurides ka tiinestumise aritmeetilist keskmist erinevates rühmades, on pea iga tunnuse puhul keskmine suurim udara tervise grupis A. Mitte ühtegi teist tunnust arvesse võttes, on tiinestumise aritmeetiline keskmine grupis A 0,532, grupis B 0,527, grupis C 0,491 ja grupis D 0,514. Seega grupis C on tiinestumisi esimesest seemendusest olnud vähim. Laktatsioonide kaupa vaadates on kindlalt näha keskmise vähenemist igas grupis alates 2. laktatsioonist. Kui 1. laktatsiooni lehmade puhul jääb tõenäosus iga grupi puhul üle 0,6, siis alates 2. laktatsioonist jääb see iga grupi puhul alla 0,5. Küll aga on tabelist 5 näha seda, et kui päevade arv poegimisest esimese seemenduseni kasvab (vahemik on 91-150), siis tiinestumise tõenäosus samuti tõuseb. Kahel korral on tiinestumine grupis D ka kõige suurema keskmisega - 3. ja suurema laktatsiooni juures ning siis, kui taastumisperioodi pikkus on 91-150 päeva. Üldiselt tuleb tabelist välja trend, kus aritmeetiline keskmine langeb grupist A grupi C-ni, aga grupi D puhul mingit kindlat jaotumist märgata pole, vaid see paikneb pigem juhuslikult.

Tabel 5: Tiinestumise tõenäosus sõltuvalt udara tervise grupist, laktatsioonist või poegimisest esimese seemenduseni kulunud päevade arvust.

Udara tervise grupp/ Tunnus	A	B	C	D	Kokku
Lakt. 1	0,634	0,625	0,625	0,627	0,632
Lakt. 2	0,475	0,479	0,459	0,456	0,477
Lakt. 3+	0,461	0,472	0,455	0,488	0,460
60-90 päeva	0,522	0,517	0,476	0,490	0,516
91-150 päeva	0,565	0,557	0,535	0,571	0,558
Kokku	0,532	0,527	0,491	0,514	0,525

Kui vaadata somaatiliste rakkude eristumise jaotumist erinevates udara tervise gruppides tihedusgraafikuna (joonis 4), siis UTG A ehk terve lehma puhul on SRE kas 0 või väga nullilähedane. Ülejäänud kolm gruppi jäävad peamiselt teisele poolele graafikust, kuid UTG D jääb peamiselt napilt alla 65 piiri, olles rohkem vasakule kaldus. Kõige suurema SRE väärtusega lehmad on grupis C, kus nende kõige suuremal määral oleva SRE väärtus on ligikaudu 90. Grupis B on SRE väärtus alates 65 joonest ühtlasemalt jaotunud, kuid kõige populaarsem väärtus jääb umbes 80 juurde.



Joonis 4: SRE jaotus udara tervise grupi kaupa.

3.3 Lehmade tiinestuvuse hindamine udara tervise grup- pide alusel

Sõltuvalt sellest, milline kontroll-lüps on võetud UTG aluseks ja kuidas on defineeritud tiinestumine, on analüüsitud:

- 93 467 sündmust, kui aluseks on esimene kontroll-lüps peale poegimist ja tiinestumise määramisel on aluseks võetud järgnev/mittejärgnev seemendus 45 päeva jooksul peale esimest seemendust (I, 45p);
- 88 360 sündmust, kui UTG on määratud esimesele seemendusele eelnenud lähima kontroll-lüpsi tulemuste põhjal ja tiinestumise määramisel on kasutatud eelnevat varianti (II, 45p);
- 87 471 sündmust, kui UTG on määratud esimesele seemendusele eelnenud kahe viimase kontroll-lüpsi halvema tulemuse kaudu ning tiinestumise määramisel on samuti kasutatud esimeses punktis mainitud viisi (III, 45p);
- 69 938 sündmust, kui aluseks on esimene kontroll-lüps peale poegimist ja tiinestumise määramisel on aluseks võetud järgneva poegimise olemasolu (I, p2);
- 66 171 sündmust, kui UTG on määratud esimesele seemendusele eelnenud lähima kontroll-lüpsi tulemuste põhjal ja tiinestumise määramisel on kasutatud eelnevat varianti (II, p2);
- 65 629 sündmust, kui UTG on määratud esimesele seemendusele eelnenud kahe viimase kontroll-lüpsi halvema tulemuse kaudu ning tiinestumise määramisel on kasutatud neljandas punktis mainitud viisi (III, p2).

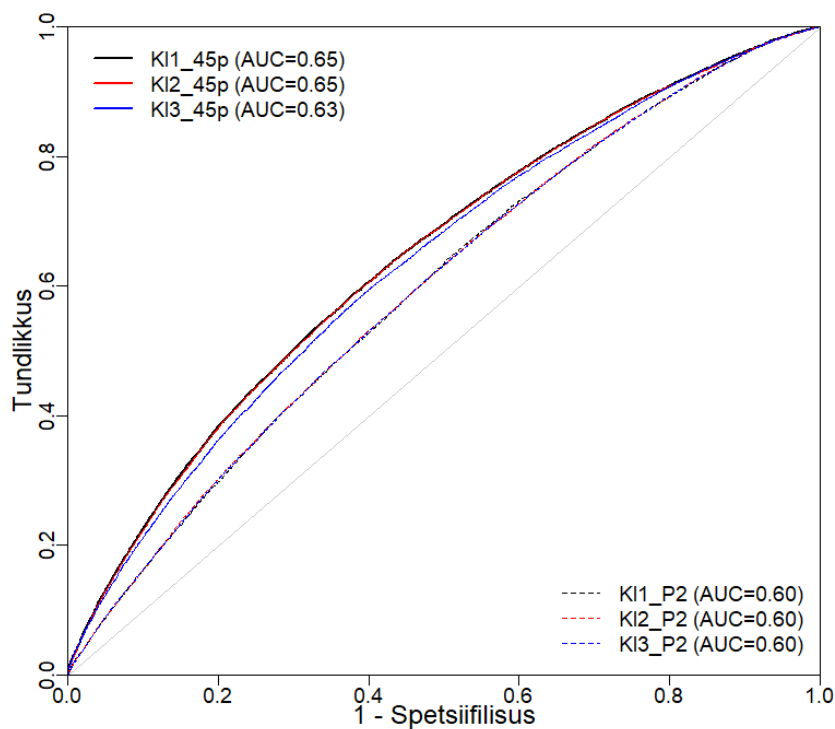
Kõikide analüüside puhul on kasutatud logit-seosefunktsiooniga üldistatud lineaarset segamudelit ehk juhusliku faktoriga logit-mudelit, kus uuritavaks tunnuseks on

tiinestumine. Argumenttunnusteks on igas mudelis faktortunnused *UTG*, *laktatsioon*, *taastumisperioodi pikkus*, *suguselekteeritud sperma kasutamine* ning vastavalt esimese kolme argumenttunnuse koosmõjud. Juhusliku faktorina on lisatud *seemendustehnik*.

Analüüsid on tehtud rakendustarkvara R abil, kasutades mudeli koostamiseks paketi *lme4* funktsiooni *glmer* (Bates *et al.*, 2015), hinnanguliste tõenäosuste leidmiseks paketti *emmeans* (Lenth, 2025) ja ROC-kõverate koostamiseks paketti *pRoc* (Robin *et al.*, 2011).

Tabelis 7 (Lisa 1) on välja toodud kõikide mudelite parameetrite hinnangud sõltumata nende statistilisest olulisusest. Nagu tabelist näha, on parameetrid mudelite kaupa küll erinevad, aga erinevus on pigem marginaalne. Mõningatel juhtudel esineb mudelite kaupa ka märgi erinevusi.

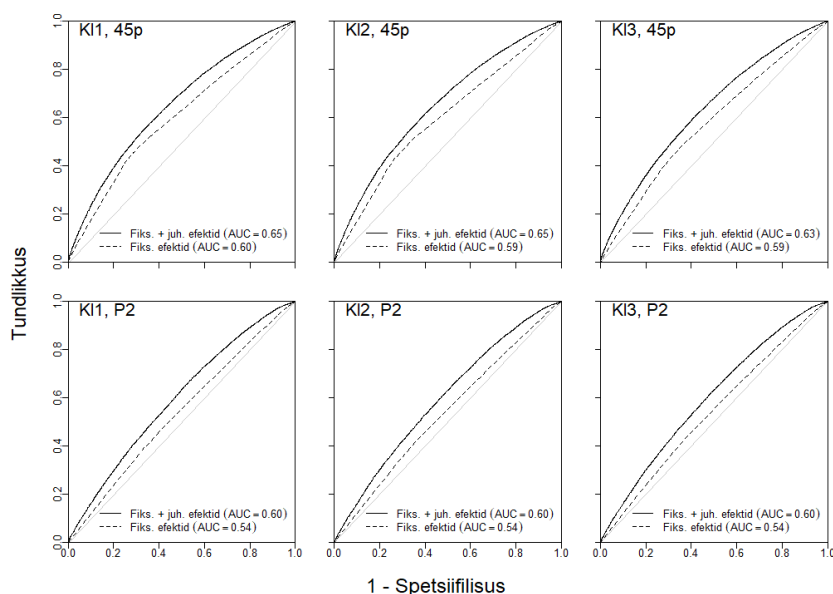
Võrreldes mudelite täpsust ROC-kõvera aluse pindala kaudu, saame, et kuuest mudelist kõige täpsemad ($AUC=0,65$) on kaks mudelit, millest mõlemal juhul on tiinestumise määramisel kasutatud 45 päeva jooksul järgneva seemenduse esinemist/mitte-esinemist ning kus ühe mudeli puhul on UTG määratud esimese kontroll-lüpsi tulemustel peale poegimist ning teise mudeli puhul esimesele seemendusele eelnenud lähima kontroll-lüpsi tulemuste abil.



Joonis 5: ROC-kõverad ja nende alused pindalad iga mudeli kohta.

Mudelid, mille andmed eeldasid järgneva poegimise olemasolu, on ebatäpsemad võrreldes ülejäänud kolme mudeliga (Joonis 5). Nende kõigi ROC-kõvera alune pindala on 0,60, mistõttu on nende mudelite täpsus väga kasin ja juhuslikkusele üpris lähedal. Ülejäänud kolm mudelit on küll kõik natukene täpsemad, kuid nende ROC-kõvera alune pindala jääb siiski alla 0,7, mis endiselt tähendab pigem ebatäpset mudelit.

Võrreldes kõigi kuue mudeli täpsust juhuga, kus ära on jäetud juhuslik faktor (seemendustehnik), on näha, et nende mudelite täpsus muutub veelgi halvemaks ja AUC on väga lähedal 0,5-le ehk sisuliselt pole nendel mudelitel ilma juhusliku faktorita mingit mõtet. (Joonis 6)



Joonis 6: ROC-kõverad iga mudeli kohta ainult fikseeritud faktorite korral vs koos juhusliku faktoriga mudelitega.

Kuna mudelid oma täpsuse poolest suuresti ei erine, siis edaspidises analüüsis on kasutusel mudel, mille puhul tiinestumine on määratud selle järgi, kas 45 päeva jooksul pärast esimest seemendust on olnud järgnev seemendus ning UTG on määratud esimesele seemendusele lähima kontroll-lüpsi SRA ja SRE põhjal. Analüüsitava mudeli valik tuli eelkõige ROC-kõvera aluse pindala suuruselt, kus see oli suurim (0,65).

Statistiliselt oluliseks osutus mudelis neli tunnust: laktatsioon, taastumisperioodi pikkus, suguselekteeritud sperma kasutamine ning laktatsiooni ja taastumisperioodi koosmõju. Udara tervise grupi faktor statistiliselt oluliseks ei osutunud ($p = 0,411$). (Lisa 3) Kuigi *UTG* tunnus tuli statistiliselt ebaoluline, on antud tunnus ja ka selle koosmõjud siiski mudelisse sisse jäetud, et täita analüüsi eesmärki uurimaks just udara tervise grupi mõju lehma tiinestumisele.

Tunnuse *UTG* analüüsimisel selgus, et võrreldes baastasemega (grupp A) on šanss tiinestuda väiksem grupis D, kus tiinestumise šanss esimesel seemendusel väheneb ligi 30%. Ka teistes gruppides on tiinestumise šansid väiksemad võrreldes grupiga

A. Grupis B on tiinestumise šanss väiksem 5,9% ja grupis C 5,0%. Olgugi et tunnus osutus ebaoluliseks, on tulemus ootuspärane, sest nakkusega lehmadel on uurimuste käigus avastatud ka muude tunnuste, näiteks piimakogus, halvenemist.

Tunnuse *Laktatsioon* puhul on võrreldes esimese laktatsiooni lehmadega tiinestumise šansid nii teise kui ka kolmanda ja vanema laktatsiooni lehmadel tunnudvalt väiksemad. Ka faktortasemetel „2” ja „3+” vahel ilmneb statistiliselt oluline erinevus, kus teise laktatsioonide lehmadel on tiinestumise šanss kõrgem veidi üle 8%.

Vaadates taastumisperioodi pikkuse mõju tiinestumisele, on vahemikuga 91-150 lehmadel üllataval kombel tiinestumise šanss kõrgem.

Kui kasutada tiinestumiseks suguselekteeritud spermat, on tiinestumine ligi 35% vähem tõenäoline. Selle põhjuseks võib olla suguselekteeritud sperma vähemate spermide arv, kui tavasperma kasutamisel ning samuti on suguselekteeritud spermide eluiga lühem, mis teeb õigeaegse seemendamise tabamise veel olulisemaks ja seetõttu suurendab tiinestumise ebaõnnestumise riski (Jaakma *et al.*, 2008).

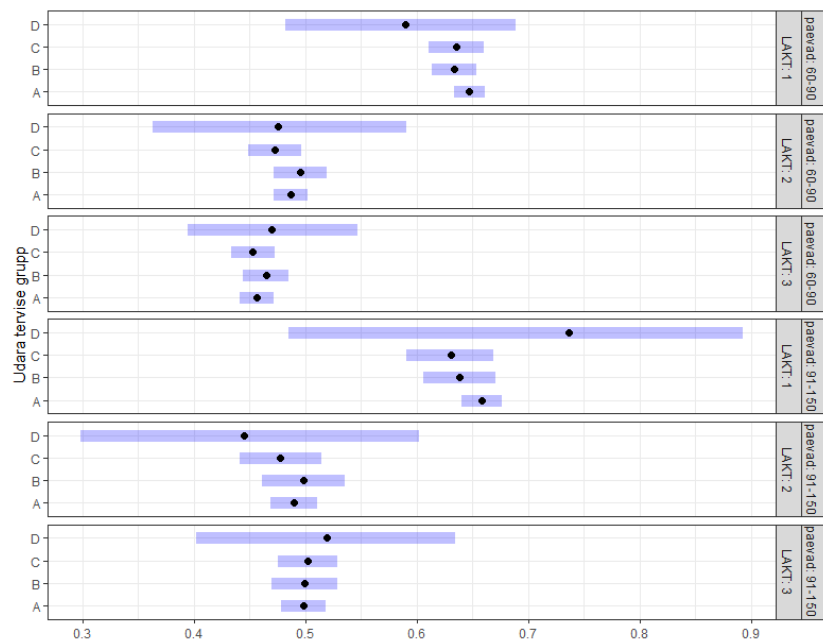
Mõnevõrra vastuolulisemad tulemused tekivad *UTG* ja *Laktatsiooni* koosmõju analüüsimisel. Seal ilmneb, et võrreldes baastasemega - *UTG A*, kus taastumisperioodi pikkus jäi vahemikku 60-90 - on udara tervise grupis D olevate kolmanda ja vanemate lehmade tiinestumise šanss ligi 1,2 korda suurem. Kuna varasemate tunnuste korral selgus, et nii *UTG-s D* kui ka 3. ja vanema laktatsiooniga lehmade tiinestumise šanss on baastasemega võrreldes väiksem, tekib siinkohal mõningane vastuolu. Aga kuna antud koosmõju ei osutunud statistiliselt oluliseks, ei saa neid tulemusi ka usaldusväärseks lugeda.

Koosmõju tunnuste *Laktatsioon* ja *Päevad* vahel osutus statistiliselt oluliseks. Võrreldes 1. laktatsiooni lehmadega, kellel aeg poegimisest esimese seemenduseni oli 60-90 päeva, on 3. ja vanemate lehmade, kelle taastumisperioodi pikkus oli 91-150 päeva, tiinestumise šanss 12,0% suurem. Teise laktatsiooni lehmadel, kellel vahemik poegimisest esimese seemenduseni oli 91-150 päeva, on tiinestumise šanss aga veidi (ca 3%) madalam.

Tabel 6: Mudeli parameetrite hinnangud eksponentfunktsiooni skaalal (mudeli parameetrite hinnangud lisas 2).

Efekt	$\exp \beta$	Efekt	$\exp \beta$
Vabaliige	1,903	C:2	0,965
UTG		D:2	1,020
B	0,944	B:3+	1,119
C	0,952	C:3+	1,016
D	0,769	D:3+	1,213
LAKT		LAKT:Päevad	
2	0,442	2:91-151	0,963
3+	0,455	3+:91-151	1,120
Päevad		UTG:LAKT:Päevad	
91-150	1,034	B:2:91-150	1,067
SS	0,743	C:2:91-150	1,203
UTG:Päevad		D:2:91-150	0,999
B:91-150	0,970	B:3+:91-150	0,978
C:91-150	0,927	C:3+:91-150	1,218
D:91-150	1,880	D:3+:91-150	0,694
UTG:LAKT			
B:2	1,049		

Joonisel 7 on kujutatud tiinestumise hinnangulised tõenäosused erinevates udara tervise gruppides sõltuvalt laktatsioonist ja taastumisperioodi pikkusest. Tõenäosuste hinnangud on leitud R-i paketi *emmeans* abil. Jooniselt on näha kindlat erinevust tiinestumise tõenäosuses sõltuvalt lehma laktatsioonist, kus esimese laktatsiooni lehmadel on tiinestumine parem. Samuti joonistub esimese laktatsiooni lehmade puhul nii vahemikus 60-90 kui ka 91-150 muster, kus tiinestumise šansid langevad vastavalt UTG halvenemisele kuni grupini C. UTG D puhul on jooniselt näha suurt usaldusvahemikku, mistõttu võib tiinestumise tõenäosus nendes kohtades palju varieeruda.



Joonis 7: Tiinestumise hinnangulised tõenäosused UTG, laktatsiooni ja taastumisperioodi koosmõjul.

Kokkuvõte

Bakalaureusetöö eesmärgiks oli uurida, kas ja kuivõrd on võimalik seemendamise eelselt toimunud kontroll-lüpsil määratud somaatiliste rakkude koguarvu ja erinevat tüüpi somaatiliste rakkude suhtarvu alusel ennustada lehmade tiinestumist esimesel seemendusel.

Analüüsi läbiviimiseks kasutati juhusliku faktoriga logistilist mudelit, kus uuritavaks tunnuseks oli lehma tiinestumine (jah/ei), fikseeritud faktortunnusteks olid lehma kontroll-lüpsi tulemustest saadud UTG, päevade arv poegimisest esimese seemenduseni ehk taastumisperioodi pikkus (kahe-tasemeline tunnus), laktatsioon (kolme-tasemeline tunnus) ja suguselekteeritud sperma kasutamine (jah/ei). Juhusliku faktori mõjuna oli mudelisse kaasatud 324 erineva tasemega seemendustehniku efekt.

Töö käigus moodustati kuus andmestikku, millest enim analüüsiti tulemusi esimesele seemendusele eelnenud lähima kontroll-lüpsi SRA ja SRE abil saadud UTG järgi, kus tiinestumine oli määratud selle põhjal, kas esimesele seemendusele järgnes 45 päeva jooksul uus seemendus või mitte. Tulemusi analüüsiti selles andmestikus enim, kuna andmete põhjal koostatud mudel oli kõige täpsem ja ülejäänud andmestikud oma sisult suuresti ei erinenud. Analüüs koostati rakendustarkvara R-is, kus mudeli koostamiseks oli kasutusel paketi *lme4* funktsioon *glmer*.

Kokku analüüsiti mainitud andmestikus 88 360 esimest seemendust. Nendest 69,9% paiknes seemendatud lehm udara tervise grupis A, 15,3% grupis B, 14,3% grupis C ja 0,5% grupis D. Keskmine tiinestumine oli 52,5%. Logistilise mudeli kohaselt mõjutasid tiinestumist statistiliselt oluliselt laktatsioon (vanemad lehmad tiinestusid vähem), taastumisperioodi pikkus (pikema taastumisperioodiga lehmad tiinestusid rohkem) ja suguselekteeritud sperma kasutamine (suguselekteeritud spermat kasutanud lehmad tiinestusid vähem).

Lisa 1

Tabel 7: Kõikide mudelite parameetrite hinnangud.

Efekt	I,45p	II,45p	III,45p	I,p2	II,p2	III,p2
Vabaliige	0,617	0,643	0,632	0,106	0.103	0.111
UTG						
B	0,011	-0,057	-0,043	-0.026	-0.017	-0.055
C	0,018	-0,049	-0,023	-0,029	-0.057	-0.025
D	-0,162	-0,263	-0,026	-0,038	-0.219	0.115
LAKT						
2	-0,654	-0,817	-0,599	-0,257	-0.261	-0.252
3+	-0,783	-0,788	-0,618	-0,312	-0.322	-0.339
Päevad						
91-150	0,057	0,033	0,028	-0.049	-0.047	-0.074
SS	-0,271	-0,297	-0,288	-0,309	-0.303	-0.301
UTG:Päevad						
B:91-150	-0,154	-0,031	0,017	-0.168	-0.040	0.063
C:91-150	0,001	-0,076	-0,054	0.009	-0.085	-0.097
D:91-150	0,132	0,631	0,124	0.114	0.034	-0.179
UTG:LAKT						
B:2	-0,046	0,048	0,096	-0.049	0.023	0.034
C:2	-0,084	-0,036	-0,064	-0.082	-0.055	-0.101
D:2	0,506	0,019	0,165	0.216	-0.030	-0.047
B:3+	-0,025	0,113	0,117	0.006	0.039	0.111
C:3+	0,023	0,016	0,012	0.051	0.066	0.049
D:3+	0,274	0,193	0,093	-0.246	0.478	-0.022
LAKT:Päevad						
2:91-150	-0,049	-0,038	0,017	0.073	0.120	0.100
3+:91-150	0,056	0,114	0,102	0.116	0.162	0.223
UTG:LAKT:Päevad						
B:2:91-150	0,164	0,065	0,028	0.332	-0.037	0.080
C:2:91-150	0,046	0,185	0,082	0.153	0.155	0.177
D:2:91-150	-0,424	-0,001	-0,226	-0.128	0.389	-0.032
B:3+:91-150	0,167	-0,023	-0,056	0.202	-0.062	-0.270
C:3+:91-150	0,140	0,198	0,132	0.173	0.091	0.054
D:3+:91-150	0,059	-0,365	-0,118	0.196	-0.480	-0.169

Lisa 2

```

> summary(T.glmr_2_45_teine)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmMod']
Family: binomial ( logit )
Formula: T ~ UTG + LAKT + S + paevad + UTG * paevad * LAKT + (1 | TECH)
Data: tiin2

      AIC      BIC    logLik -2*log(L)  df.resid
117049.6 117293.8 -58498.8 116997.6    88334

Scaled residuals:
   Min       1Q   Median       3Q      Max
-2.8615 -0.9406  0.5532  0.9353  2.1772

Random effects:
 Groups Name      Variance Std.Dev.
  TECH (Intercept) 0.1898   0.4357
Number of obs: 88360, groups:  TECH, 323

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.643239   0.032236  19.954 <2e-16 ***
UTGB           -0.057138   0.037860  -1.509  0.1312
UTGC           -0.048917   0.049710  -0.984  0.3251
UTGD           -0.262987   0.212015  -1.240  0.2148
LAKT2          -0.817114   0.023088 -35.392 <2e-16 ***
LAKT3          -0.788286   0.022682 -34.754 <2e-16 ***
S              -0.296778   0.032548  -9.118 <2e-16 ***
paevad91-150   0.033462   0.034951   0.957  0.3384
UTGB:paevad91-150 -0.030933   0.081715  -0.379  0.7050
UTGC:paevad91-150 -0.076318   0.099900  -0.764  0.4449
UTGD:paevad91-150  0.631085   0.494798   1.275  0.2022
UTGB:LAKT2     0.048113   0.058079   0.828  0.4074
UTGC:LAKT2    -0.036052   0.067235  -0.536  0.5918
UTGD:LAKT2     0.019461   0.319451   0.061  0.9514
UTGB:LAKT3     0.112762   0.052086   2.165  0.0304 *
UTGC:LAKT3     0.016063   0.059933   0.268  0.7887
UTGD:LAKT3     0.193388   0.259198   0.746  0.4556
LAKT2:paevad91-150 -0.037724   0.050462  -0.748  0.4547
LAKT3:paevad91-150  0.113651   0.048854   2.326  0.0200 *
UTGB:LAKT2:paevad91-150 0.065223   0.120627   0.541  0.5887
UTGC:LAKT2:paevad91-150 0.184516   0.133378   1.383  0.1665
UTGD:LAKT2:paevad91-150 -0.001352   0.646524  -0.002  0.9983
UTGB:LAKT3:paevad91-150 -0.022725   0.108915  -0.209  0.8347
UTGC:LAKT3:paevad91-150 0.197561   0.119672   1.651  0.0988 .
UTGD:LAKT3:paevad91-150 -0.365261   0.556745  -0.656  0.5118

```

Joonis 8: II, 45p mudeli parameetrite hinnangud koos muude näitajatega RStudio programmis.

Lisa 3

```
> Anova(T.glmr_2_45_teine)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: T

          Chisq Df Pr(>Chisq)
UTG          2.8796  3  0.4105690
LAKT        2670.6752  2 < 2.2e-16 ***
S           83.1400  1 < 2.2e-16 ***
paevad       19.9875  1  7.795e-06 ***
UTG:paevad    7.2608  3  0.0640327 .
UTG:LAKT      6.8051  6  0.3392436
LAKT:paevad   15.3336  2  0.0004681 ***
UTG:LAKT:paevad 4.5004  6  0.6092819
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Joonis 9: II, 45p mudeli tunnuste statistilised olulisused RStudio programmis.

Kasutatud allikad

- Agresti, A. (2002). *Categorical Data Analysis*. 2. väljaanne. A John Wiley & Sons, lk. 165–180, 492–516.
- Bates, D., M. Mächler, B. Bolker ja S. Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. *Journal of Statistical Software* 67.1, lk. 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- EPJ, Eesti Põllumajandusloomade Jõudluskontroll (i.a). *SRE - soomaatiliste rakkude eristamine*. Vaadatud 2025-04-04. URL: <https://www.epj.ee/jkk/piimaveised/piimaveiste-j%C3%B5udluskontrolli-lisateenused/sre.html>.
- EPJ, Eesti Põllumajandusloomade Jõudluskontroll AS (2024). *Eesti jõudluskontrolli aastaraamat*, lk. 16.
- ICBF, Irish Cattle Breeding Federation (2012). *All about Somatic Cell Counts*. URL: <https://www.icbf.com/health-and-disease-home/all-about-somatic-cell-counts/>.
- Jaakma, Ü., M. Jalakas, J. Kurõkin ja L. Majas (2008). *Suguselekteeritud sperma kasutamine piimakarjakasvatustes*. URL: <https://www.epj.ee/jkk/piimaveised/piimaveiste-j%C3%B5udluskontrolli-kasulikteave/suguselekteeritud-sperma-kasutamine-piimakarjakasvatustes-august-2008.html>.
- Kaart, T. (2012). *Binaarsete tunnuste analüüsimeetodid*. Vaadatud 2025-04-23. URL: http://ph.emu.ee/~ktanel/bin_tunnuste_analyys/.
- Kalmus, P. (2019). *Madalam SRA piirarv*. URL: [https://www.epj.ee/jkk/piimaveised/piimaveiste-j%C3%B5udluskontrolli-kasulikteave/madalam-sra-piirarv-\(jaanuar-2019\).html](https://www.epj.ee/jkk/piimaveised/piimaveiste-j%C3%B5udluskontrolli-kasulikteave/madalam-sra-piirarv-(jaanuar-2019).html).

- Lenth, Russell V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Vaadatud 2025-05-05. URL: <https://CRAN.R-project.org/package=emmeans>.
- Lillik, M., A. Pentjärv, L. Taaler ja K. Ilves (2021). *Somaatiliste rakkude eristamine - mida oleme aasta jooksul teada saanud?* Vaadatud 2025-04-02. URL: <https://dea.digar.ee/cgi-bin/dea?a=d&d=jksonumidpiimaveise20211101.2.8&e=-----et-25--1--txt-txIN%7ctxTI%7ctxAU%7ctxTA----->.
- McCulloch, C. E. ja S. R. Searle (2001). *Generalized, Linear and Mixed Models*. A John Wiley & Sons, lk. 135–246.
- Nahm, F. S. (2022). “Receiver operating characteristic curve: overview and practical use for clinicians.” *Korean Journal of Anesthesiology*. DOI: <https://doi.org/10.4097/kja.21209>.
- Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez ja M. Müller (2011). “pROC: an open-source package for R and S+ to analyze and compare ROC curves”. *BMC Bioinformatics*. DOI: <https://doi.org/10.1186/1471-2105-12-77>.
- Schwarz, D., S. Kleinhans, G. Reimann, P. Stückler, F. Reith, K. Ilves, K. Pedastaar, L. Yan, Z. Zhang, M. Valdivieso, M. L. Barreal ja R. Fouz (2020). “Investigation of dairy cow performance in different udder health groups defined based on a combination of somatic cell count and differential somatic cell count”. *Preventive Veterinary Medicine*. DOI: <https://doi.org/10.1016/j.prevetmed.2020.105123>.
- Tooding, L. M. (2007). *Andmete analüüs ja tõlgendamine sotsiaalteadustes*. Tartu Ülikooli Kirjastus, lk. 281–288.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Karoliine Inselberg,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Eesti holsteini tõugu piimalehmade tiinestumine hinnatuna nende piima erinevat tüüpi somaatiliste rakkude suhtarvu ja koguarvu alusel”, 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.

Karoliine Inselberg

13.05.2025