

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Mari-Ann Liidemaa
**Vastavushindamine metroloogias: Bayesi ja
klassikalise statistika võrdlev käsitus**

Matemaatiline statistika
Bakalaureusetöö (9 EAP)

Juhendajad: DSc Aigar Vaigu,
PhD Märt Möls

TARTU 2026

VASTAVUSHINDAMINE METROLOOGIAS: BAYESI JA KLASSIKALISE STATISTIKA VÕRDLEV KÄSITLUS

Bakalaureusetöö

Mari-Ann Liidemaa

Lühikokkuvõte

Käesoleva bakalaureusetöö eesmärk on võrrelda klassikalist ja Bayesi lähenemist mõõtemääramatuse ning vastavushindamise kontekstis. Teoreetilises osas käsitletakse mõõtemääramatust, tõenäosustõlgendusi ning vastandatakse *TOST*-testi ja Bayesi järeldaotusel põhinevat hindamist.

Praktilises osas analüüsiti Metroserdi nädisandmeid 500-grammise testvihi massivea kohta. Nii klassikaline kui ka Bayesi analüüs kinnitasid, et kuigi erinevus etalonist on statistiliselt oluline, jääb viga lubatud piiridesse ja viht vastab nõuetele. Simulatsiooni abil võrreldi meetodite käitumist A–B–B–A mõõtetsükli korral, mis näitas, et vaadeldud tingimustes on tulemuste erinevus Bayesi ja klassikalise lähenemise vahel äärmiselt väike. Väikese mõõtemüra ja piisavalt suure valimi tõttu domineerib Bayesi analüüsis tõepärafunktsioon ning eeljaotuse roll on väike.

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

Märksõnad: Bayesi statistika, mõõtemääramatus, vastavushindamine, metroloogia, ekvivalentsustest.

CONFORMITY ASSESSMENT IN METROLOGY: A COMPARATIVE TREATMENT OF BAYESIAN AND CLASSICAL STATISTICS

Bachelor thesis

Mari-Ann Liidemaa

Abstract

The aim of this bachelor's thesis is to compare classical and Bayesian statistical approaches in the context of measurement uncertainty and conformity assessment. The theoretical part addresses measurement uncertainty and interpretations of probability, contrasting the *TOST* (Two One-Sided Tests) procedure with Bayesian assessment based on the posterior distribution.

In the practical part, sample data from Metrosert regarding the mass error of a 500 g weight were analyzed. Both classical and Bayesian analyses confirmed that although the difference from the standard is statistically significant, the error remains within the specified tolerance limits and the weight conforms to requirements. A simulation was conducted to compare the behavior of both methods during an A–B–B–A measurement cycle, demonstrating that the results are very similar under the observed conditions. Due to low measurement noise and a sufficiently large sample size, the likelihood function dominates the Bayesian analysis, and the role of the prior distribution is minimal.

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

Key Words: Bayesian statistics, measurement uncertainty, conformity assessment, metrology, equivalence testing.

Sisukord

Sissejuhatus	4
1 Mõõtemääramatus mõõteteaduses	5
2 Tõenäosuse interpretatsioonid	8
2.1 Sageduslik statistika	8
2.2 Bayesi statistika	9
2.2.1 Tõenäosusvahemik	11
2.2.2 Punkthinnang	11
2.2.3 Hüpoteeside testimine	11
2.3 Bayesi ja sagedusliku lähenemise näide	13
3 Vastavushindamine	16
3.1 Klassikaline statistiline käsitlus	17
3.2 Bayesi käsitlus vastavushindamisel	19
4 Andmeanalüüs	20
4.1 Andmestiku ülevaade	20
4.2 TOST-test ekvivalentsuse hindamisel	23
4.3 Bayesi analüüs	25
4.4 A–B–B–A mõõtetsükli simulatsioon	29
4.4.1 Simulatsiooni ülesehitus	30
4.4.2 Hindamiskriteeriumid	30
4.4.3 Simulatsiooni tulemused	32
Kokkuvõte	34
Kasutatud allikad	35

Sissejuhatus

Mõõtemääramatuse¹ mõiste on mõõtmisajaloos suhteliselt uus nähtus, kuigi viga ja veaanalüüs on omanud tähtsat rolli mõõteteaduse ehk metroloogia valdkonnas juba pikka aega. Isegi parandatud tulemuse õigsuse suhtes jääb alati üles määramatus ehk kahtlus selles, kui hästi mõõtmistulemus esindab mõõdetava suuruse tegelikku väärtust. (JCGM, 2008)

Mõõtemääramatus muutub eriti oluliseks vastavushindamisel, kus mõõtetulemuse põhjal otsustatakse, kas uuritav objekt vastab etteantud nõuetele. Metroloogias tähendab see sageli kontrollimist, kas mõõtevahendi või etaloni tegelik väärtus jääb lubatud tolerantsipiiridesse. Kuna mõõtetulemus ei lange üldjuhul täpselt kokku mõõtesuuruse tegeliku väärtusega, kaasneb otsusega alati teatud risk: nõuetele mittevastav objekt võidakse ekslikult vastavaks lugeda või vastupidi. (JCGM, 2012)

Käesoleva bakalaureusetöö eesmärk on võrrelda klassikalist ja Bayesi lähenemist mõõtemääramatuse ning vastavushindamise kontekstis. Töös antakse ülevaade mõõtemääramatuse hindamisest metroloogias, tõenäosuse sageduslikust ja Bayesi tõlgendusest ning vastavushindamise statistilistest põhimõtetest. Klassikalise käsitlusena vaadeldakse ekvivalentsuse testimist kahe ühepoolse testi ehk *TOST*-testi abil, Bayesi käsitluses hinnatakse vastavust järeljaotuse põhjal.

Töö praktilises osas analüüsitakse Eesti metroloogia keskasutuse näidisandmes-
tikku. Analüüsis rakendatakse ekvivalentsuse hindamiseks *TOST*-testi ning Bayesi meetodeid. Lisaks võrreldakse mõlema meetodi käitumist simulatsioonide abil. Analüüsimiseks kasutatakse statistikatarkvara R.

¹Töös kasutatakse mõisteid määramatus ja mõõtemääramatus sünonüümidena.

1 Mõõtemääramatus mõõteteaduses

Järgnev peatükk põhineb teosel „Evaluation of measurement data — Guide to the expression of uncertainty in measurement“ (JCGM, 2008).

Selleks et hinnata mõõtemääramatust, tuleb esmalt defineerida mõõtemudel. Enamasti ei mõõdetata mõõtesuurust Y otse, vaid see saadakse suuruste X_i , $i = 1, 2, \dots, N$ abil seose f kaudu:

$$Y = f(X_1, X_2, \dots, X_N). \quad (1)$$

Sisendid $X_1, X_2, \dots, X_N$ võivad olla nii konstandid, parandid kui ka mõõdetavad, millede hinnangud jagatakse hindamismeetodi alusel kaheks:

1. suurused, mille väärtused ja määramatused saadakse kordusmõõtmiste käigus;
2. suurused, mille algandmed pärinevad välistest allikatest, näiteks kalibreeritud mõõteetalonidega kaasatulevast dokumentatsioonist.

Näiteks võib kalibreerimistunnistusel olla märgitud, et ühekilogrammise etalonmassi hinnanguline väärtus on 1000,000325 g ning mõõtemääramatus 240 μg . Standardmääramatus saadakse sellisel juhul määramatuse jagamisel vastava kateteguriga, mille tähendust käsitletakse peatüki lõpuosas

$$u(m_S) = \frac{240 \mu\text{g}}{3}.$$

Seega ei põhine antud määramatus kordusmõõtmistel, vaid eelneval teadmisel mõõtesuuruse võimaliku hajuvuse kohta.

Määramatuse kvantifitseerimiseks kasutatakse kahte hindamismetoodikat: A-tüüpi ja B-tüüpi hindamine. A-tüüpi hindamine tugineb vaatlustulemuste statistilisele

analüüsile, kus määramatust iseloomustatakse dispersioonide s_i^2 või standardhälvetele s_i ning vabadusastmete arvu kaudu. B-tüüpi hindamise puhul leitakse dispersioonide ligikaudsed hinnangud u_j^2 muude meetodite abil.

Kui tunnused on omavahel korreleeritud, tuleb hinnata nende vahelisi kovariatsioone $u(x_i, x_j)$. Seose tugevust iseloomustab korrelatsioonikordaja:

$$r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)},$$

kus $r(x_i, x_j) = r(x_j, x_i)$ ja $-1 \leq r(x_i, x_j) \leq 1$. Kui x_i ja x_j on sõltumatud, siis $r(x_i, x_j) = 0$, kuid vastupidine ei pruugi kehtida. Nullkorrelatsioonist ei järeldu sõltumatus, sest korrelatsioon mõõdab ainult lineaarset seost.

Mõõdetava suuruse Y hinnang y ehk mõõtetulemus saadakse valemist (1), kasutades hinnanguid $x_1, x_2, \dots, x_N$ suuruste $X_1, X_2, \dots, X_N$ väärtustele. Seega saab hinnangu y valemiga

$$y = f(x_1, x_2, \dots, x_N).$$

Lõplik mõõtetulemus tuleb esitada liitmõõtemääramatusega $u_c(y)$, mis on saadud A- ja B-tüüpi määramatuste liitmisel:

$$u_c^2(y) = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j),$$

kus x_i ja x_j on suuruste X_i ja X_j hinnangud ning $u(x_i, x_j)$ kovariatsiooni hinnang. Osatuletiste $\frac{\partial f}{\partial x_i}$ all mõeldakse siin funktsiooni osatuletiste väärtusi punktis $(x_1, x_2, \dots, x_N)$. Kui sisendsuurused on sõltumatud, on mõõtetulemuse liitdispersioon kujul

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i).$$

Võttes näiteks funktsiooni $y = 2x_1 + 2x_2$, kus x_1 ja x_2 on sõltumatud, siis

$$\left(\frac{\partial f}{\partial x_1} \right)^2 = 2^2 = 4 \quad \text{ja} \quad \left(\frac{\partial f}{\partial x_2} \right)^2 = 2^2 = 4.$$

Seega $u_c(y) = \sqrt{4u^2(x_1) + 4u^2(x_2)}$.

Liitmõõtemääramatuse $u_c(y)$ kaudu saab tuletada laiendmääramatuse U , et määrata usaldusvahemik $(y - U, y + U)$, mis teatud kõrgema usaldatavusega hõlmab tegeliku mõõtesuuruse väärtuse. Laiendmääramatuse saamiseks korrutatakse liitmõõtemääramatus $u_c(y)$ läbi katteteguriga k :

$$U = k \cdot u_c(y).$$

Katteteguri k valik sõltub soovitud usaldatavuse tasemest p . Kui juhusliku suuruse Y jaotuseks on normaaljaotus, siis valides katteteguri väärtuseks $k = 2$, saadakse ligikaudu 95%-line usaldusvahemik mõõdetava suuruse tegelikule väärtusele. Olukordades, kus informatsiooni on vähem, võib osutuda vajalikuks kasutada suuremat kattetegurit, tuginedes t -jaotusele ja efektiivsele vabadusastmete arvule.

Seega esitatakse mõõtetulemus kujul $Y = y \pm U$. Samaväärselt võib tulemust väljendada kui $y - U \leq Y \leq y + U$.

2 Tõenäosuse interpretatsioonid

Bayesi statistika erineb sageduslikust selle poolest, et kui esimeses vaadeldakse mõõdetavat suurust kui juhuslikku suurust mingisuguse kindla jaotusega, siis sageduslikus statistikas on mõõdetav suurus tundmatu, kuid fikseeritud, ja juhuslikkus tuleneb mõõtmisest. Bayesi lähenemine võimaldab lisaks vaatlusandmetele võtta arvesse ka eelnevaid teadmisi, mistõttu on see eriti kasulik olukordades, kus andmeid on vähe ja ainult valimil põhinevad hinnangud võivad olla ebatäpsed.

Järgnevad alapeatükid on kirjutatud „Disentangling Classical and Bayesian Approaches to Uncertainty Analysis“ põhjal, kui ei ole viidatud teisiti (Willink ja White, [s.a.](#)).

2.1 Sageduslik statistika

Sageduslikus statistikas tõlgendatakse tõenäosust sündmuse esinemissagedusena korduvate katsete korral. Sageduslikus statistikas on kesksel kohal usaldusvahemik ehk juhuslik vahemik, millel on kindlaksmääratud tõenäosus sisaldada hinnatavat parameetrit.

Boltzmanni konstandi k nihketa mõõtmisel eeldatakse, et mõõtevead on normaaljaotusega. Seega on k mõõtmistulemuste populatsiooni keskväärus ning mõõtmise eesmärk on seda hinnata.

Olgu n sõltumatut mõõtmist, eeldades, et vaatlused pärinevad normaaljaotusest keskväärusega k ja standardhälbega σ . Tähistagu $\bar{X}$ valimi keskmist ning S standardhälvet. Sel juhul sisaldab juhuslik intervall

$$\left(\bar{X} - t_{0,975;n-1} \frac{S}{\sqrt{n}}; \bar{X} + t_{0,975;n-1} \frac{S}{\sqrt{n}} \right) \quad (2)$$

parameetrit k tõenäosusega 0,95, kus $t_{0,975;n-1}$ tähistab vabadusastmega $n - 1$

vastava t -jaotuse 0,975-kvantiili. Intervalli (2) nimetatakse k 95%-liseks usaldusvahemikuks. Andmete põhjal saab arvutada hinnangud $\bar{x}$ ja s , mis vastavad juhuslike suuruste $\bar{X}$ ja S realiseerunud väärtustele, kusjuures nendele vastav 95%-line usaldusvahemik on

$$\left(\bar{x} - t_{0,975;n-1} \frac{s}{\sqrt{n}}; \bar{x} + t_{0,975;n-1} \frac{s}{\sqrt{n}} \right).$$

Enne mõõtmiste tegemist on tõenäosus, et saadud juhuslik vahemik (2) sisaldab k -d, 0,95. Pärast mõõtmiste tegemist ei ole usaldusintervall enam juhuslik ning kas sisaldab või ei sisalda tegelikku parameetri väärtust k . Sagedusliku tõlgenduse kohaselt tähendab 95%-line usaldusintervall seda, et kui sama mõõtmisprotseduurri korrata suurel arvul kordadel, siis ligikaudu 95% selliselt saadud intervallidest sisaldavad parameetri k tegelikku väärtust.

2.2 Bayesi statistika

Bayesi lähenemisviisis kirjeldatakse teadmisi parameetri kohta tõenäosusjaotuse abil, mida pidevalt andmete põhjal täiendatakse. Pidevate jaotuste korral saab Bayesi teoreemi

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

kirjutada kujul

$$P(A|B) = \frac{P(B|A)P(A)}{\int P(B|A)P(A) dA}, \quad (3)$$

kus nimetaja on normaliseerimiskonstant, et järeljaotuse alune pindala oleks 1 ja funktsioon rahuldaks tihedusfunktsiooni tingimusi, ning

- $P(A)$ on eeljaotus (ingl *prior*) ehk teadmised parameetri A kohta enne andmete vaatamist;
- $P(B|A)$ on tõepärafunktsioon (ingl *likelihood*), mis näitab tõenäosust saada mõõtmistulemused B antud parameetri A väärtuse korral;
- $P(A|B)$ on järeljaotus (ingl *posterior*) ehk teadmised parameetri A kohta pärast andmete vaatamist.

Seega saab valemi (3) kirjutada lihtsustatud kujul

$$\text{järeljaotus} = \frac{\text{tõepärafunktsioon} \times \text{eeljaotus}}{\text{normeeriv konstant}}. \quad (4)$$

Vaadeldakse taas Boltzmanni konstandi k mõõtmist juhul, kui σ on teada. Olgu meil n vaatlust, mille korral valimi keskmise juhuslik suurus on $\bar{X}$ ning selle realiseerinud väärtus $\bar{x}$. Olgu $\bar{X}$ normaaljaotusega juhuslik suurus keskväärtusega k ja dispersiooniga $\frac{\sigma^2}{n}$. Sel juhul saab järeljaotuse esitada kujul

$$p_{\text{post}}(k = z) = \frac{p(\bar{X} = \bar{x} | k = z) p_{\text{prior}}(k = z)}{\int p(\bar{X} = \bar{x} | k = z) p_{\text{prior}}(k = z) dz},$$

kus $p(\bar{X} = \bar{x} | k = z)$ on tõepärafunktsioon, mis kirjeldab tõenäosustihedust, et juhuslik muutuja $\bar{X}$ omandab väärtuse $\bar{x}$ tingimusel $k = z$. Funktsioonid $p_{\text{prior}}(k = z)$ ja $p_{\text{post}}(k = z)$ tähistavad vastavalt parameetri k eel- ja järeljaotuse tihedusfunktsioone. Täpsemalt on p_{post} parameetri väärtuse $k = z$ tinglik tõenäosus antud andmete korral.

Eeljaotuse tihedusfunktsioon $p_{\text{prior}}(k = z)$ väljendab eksperdi eelnevaid teadmisi või uskumusi parameetri k kohta. Kui eeljaotuse tihedusfunktsioon on normaaljaotusega, mille keskväärtus on μ_{prior} ja dispersioon σ_{prior}^2 , siis on ka k järeljaotus normaaljaotusega parameetritega

$$\mu_{post} = \frac{\mu_{prior} \frac{\sigma^2}{n} + \bar{x} \sigma_{prior}^2}{\frac{\sigma^2}{n} + \sigma_{prior}^2} \quad \text{ja} \quad \sigma_{post}^2 = \frac{\frac{\sigma^2}{n} \sigma_{prior}^2}{\frac{\sigma^2}{n} + \sigma_{prior}^2}.$$

Seega on järeljaotuse keskvääratus eeljaotuse keskvääratuse ja valimi keskmise kaalutud keskmine. Lisaks on järeljaotuse dispersioon väiksem kui klassikaline dispersioon $\frac{\sigma^2}{n}$, kuna analüüsi on kaasatud täiendav eelnev informatsioon.

2.2.1 Tõenäosusvahemik

Kui järeljaotus on normaaljaotusega keskvääratusega μ_{post} ja standardhälbega σ_{post} , saab leida k 95%-lise tõenäosusvahemiku

$$(\mu_{post} - 1,96\sigma_{post}; \mu_{post} + 1,96\sigma_{post}).$$

Erinevalt sageduslikust statistikast võib Bayesi lähenemise puhul pärast andmete vaatlemist väita, et 95%-lise tõenäosusega jääb k sellesse vahemikku.

2.2.2 Punkthinnang

Praktikas on uuritava parameetri kirjeldamiseks sageli vaja esitada üks konkreetne arvuline hinnang ehk punkthinnang, milleks kasutatakse järeljaotust iseloomustavaid näitajaid, nagu keskvääratus, mood (ingl *Maximum a Posteriori*, lühemalt MAP) või mediaan (Bolstad ja Curran, 2017). Käesolevas töös kasutatakse punkthinnanguna järeljaotuse keskvääratust, kuna vaadeldavate normaaljaotuste korral langevad keskvääratus, mediaan ja mood kokku ning seetõttu ei ole nende karakteristikute vahel sisulist erinevust.

2.2.3 Hüpoteeside testimine

Kui soovitakse testida ühepoolset hüpoteesipaari, sõnastatakse hüpoteesid

$$H_0 : \theta \leq \theta_0 \quad \text{ja} \quad H_1 : \theta > \theta_0,$$

kus θ tähistab sündmuse toimumise tõenäosust. Integreerides järeljaotuse tihedusfunktsiooni üle vastava piirkonna, saab arvutada järeljaotusele vastava nullhüpoteesi tõenäosuse (ingl *post-data probability*)

$$P(H_0 : \theta \leq \theta_0 | \text{andmed}) = \int_0^{\theta_0} p_{\text{post}}(\theta) d\theta.$$

Seega testitakse Bayesi lähenemise puhul hüpoteese jällegi parameetri järeljaotuse abil ning nullhüpotees kummutatakse juhul, kui tõenäosus on väiksem olulisuse nivoost α . (Bolstad ja Curran, 2017)

Punkthüpoteesiga kirjeldatud täpse väärtuse tõenäosus on null, mistõttu ei saa hüpoteesi testimiseks järeljaotuse tõenäosust kasutada. Selle asemel leitakse parameetri θ jaoks tõenäosusvahemik $(1 - \alpha) \cdot 100\%$. Kui θ_0 asub selles tõenäosusvahemikus, siis jäädakse nullhüpoteesi juurde. (Bolstad ja Curran, 2017)

Lisaks järeljaotuse tõenäosusele kasutatakse Bayesi hüpoteeside testimisel sageli ka Bayesi faktorit (lühemalt BF), mis avaldub kujul

$$\text{BF} = \frac{P(\text{andmed} | H_1)}{P(\text{andmed} | H_0)}.$$

Kui enne andmete vaatlemist eeldada, et null- ja alternatiivne hüpotees kehtivad mõlemad 50%-lise tõenäosusega, st $P(H_0) = P(H_1) = 0,5$, siis avaldub Bayesi teoreemi põhjal Bayesi faktor hüpoteeside järeljaotuste tõenäosuste suhtena

$$\frac{P(\text{andmed} | H_1)}{P(\text{andmed} | H_0)} = \frac{P(H_1 | \text{andmed}) / P(H_1)}{P(H_0 | \text{andmed}) / P(H_0)} = \frac{P(H_1 | \text{andmed})}{P(H_0 | \text{andmed})}.$$

Sellisel juhul saab Bayesi faktori abil järeljaotusele vastava nullhüpoteesi tõenäo-

suse arvutada kujul

$$P(H_0|\text{andmed}) = \frac{1}{1 + \text{BF}}.$$

Bayesi faktorit väärtusega $\text{BF} = 20$ tõlgendatakse kui tugevat kinnitust alternatiivse hüpoteesi kehtimisele (Wei *et al.*, 2022). Näiteks juhul, kui $\text{BF} = 20$ ja $P(H_0) = P(H_1) = 0,5$, siis

$$P(H_0|\text{andmed}) = \frac{1}{1 + \text{BF}} = \frac{1}{21} \approx 0,0476,$$

mis tähendab, et pärast andmete vaatlemist on nullhüpoteesi kehtimise tõenäosus väiksem kui 5%. Seega juhul, kui eeljaotus on mitteinformatiivne, vastab $\text{BF} = 20$ olukorrale, kus nullhüpoteesi järeljaotuse tõenäosus on väiksem kui 5%.

2.3 Bayesi ja sagedusliku lähenemise näide

Südameinfarkti põdevatele patsientidele on tavapärane manustada ravimeid, mis lahustavad verehüübeid. Uuringusse kaasatud patsientide andmed on saadud Šotimaal ajavahemikus detsember 1988 kuni detsember 1991 läbi viidud randomiseeritud uuringu tulemusena. Uuringu eesmärk oli mõõta suremuse šansi kahanemist juhul, kui trombolüütilist preparaati (anistreplaasi) manustatakse patsiendile kohe kodus, mitte alles haiglasse jõudes. (Possolo, 2015)

Tabel 1: Eksperimendi tulemused.

Grupp		Kodus	Haiglas
Tulemus	Suri	13	23
	Jäi ellu	150	125

Tabelist on näha, et $\hat{\theta} = \log(\hat{OR}) = \log\left(\frac{13 \cdot 125}{23 \cdot 150}\right) \approx -0,75$ ja $\hat{\sigma}^2 = \frac{1}{13} + \frac{1}{23} + \frac{1}{150} + \frac{1}{125} \approx 0,14$. Seega kuna $\hat{\theta} < 0$, vähendab varajane ravi surma šanssi mär-

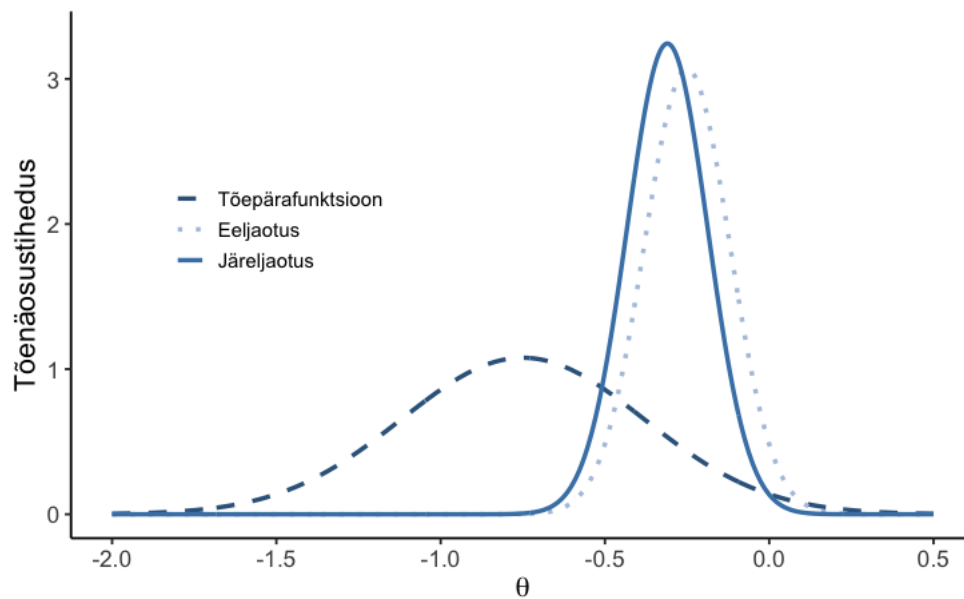
kimisväärselt (53%) võrreldes hilisema raviga. Vastav 95%-line usaldusintervall $\log(OR)$ -ile tuli $(-1,48; -0,02)$. (Possolo, 2015)

On tõenäoline, et anistreplaasi varajane manustamine ei oleks näidanud sellist mõju, kui uuringusse oleks kaasatud rohkem patsiente. Sellest skeptitsismist lähtudes viidi läbi alternatiivne Bayesi analüüs, mis arvestab kogenud kardioloogi eksperthinnangut ning kasutab eeljaotusena normaaljaotust keskväärtusega $\mu_{prior} = -0,26$ ja standardhälbega $\sigma_{prior} = 0,13$. Eksperdi arvamuse kohaselt on 15 – 20% suremuse kahanemine usutav, kuid samas on vähetõenäoline nii igasuguse kasu puudumine kui ka nii drastiline (üle 40%) paranemine, nagu antud andmed viitaksid. (Possolo, 2015)

Rakendades Bayesi meetodit koos vastava eeljaotuse ning tõepärafunktsiooniga, saadakse θ jaoks järeljaotus, mille keskväärtus on

$$\frac{\frac{\theta}{\sigma^2} + \frac{\mu_{prior}}{\sigma_{prior}^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_{prior}^2}} = -0,31,$$

kus θ tähistab šansside suhte logaritmi ($\theta = \log(OR)$), μ_{prior} eeljaotuse keskväärtust ja σ^2 parameetri θ hinnangu dispersiooni. Seega vähendab varajane ravi suremuse šanssi vaid $1 - e^{-0,31} = 27\%$. (Possolo, 2015)

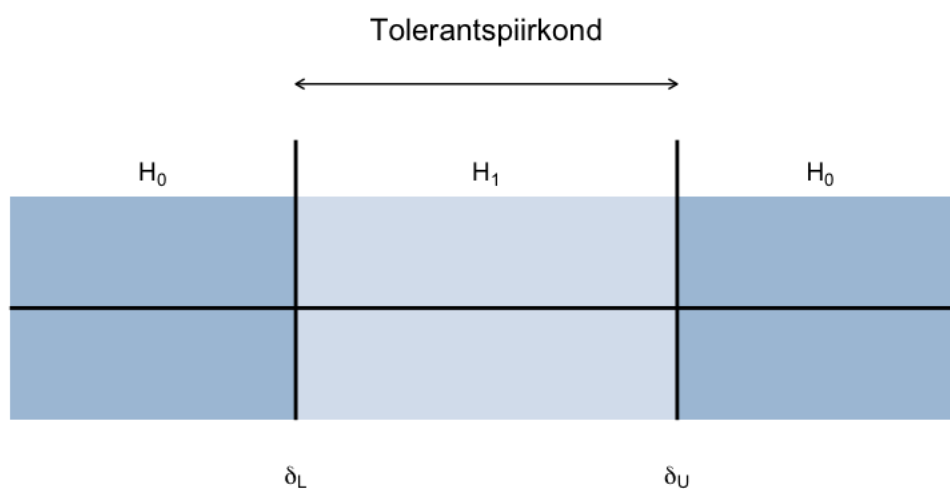


Joonis 1: Parameetri $\theta = \log(OR)$ eeljaotus, tõepärafunktsioon ja järeljaotus.

Joonisel 1 on kujutatud eeljaotuse tihedusfunktsioon, tõepärafunktsioon ning järeljaotuse tihedusfunktsioon, millest on näha, et andmed nihutavad eeljaotust vaid vähesel määral vasakule ehk väiksemate θ väärtuste suunas. Kuna järeljaotus on normaaljaotus keskvaartusega $-0,31$ ja standardhälbega $0,12$, on 95%-line tõenäosusvahemik suremuse kahanemisele ligikaudu 7% kuni 42%. (Possolo, 2015)

3 Vastavushindamine

Vastavushindamise käigus tehakse kindlaks, kas objekt (näiteks mõõtevahend) vastab seatud nõuetele. Metrooloogilises kontekstis väljendatakse neid nõudeid tavaliselt tolerantsipiirkonnana, mis määratleb lubatud väärtuste vahemiku, kuhu mõõtesuuruse tegelik väärtus jääma peab. Sõltuvalt spetsifikatsioonist võib see olla piiratud kas ühe või kahe tolerantsipiiriga, näiteks võib olla määratud vaid üks ülemine piir, millest mõõdetav ei tohi suurem olla. Nende piiride põhjal määratakse hüpoteesipiirkonnad, mida on kujutatud joonisel 2. (JCGM, 2012)



Joonis 2: Hüpoteeside piirkonnad.

Vastavushindamise eesmärk on näidata, et mõõtevahend vastab nõuetele ehk tegelik erinevus jääb tolerantsipiiridesse. Seetõttu sõnastatakse alternatiivne hüpotees ekvivalentsuse kaudu ning mõõtmiste vähesuse või ebakindluse korral jäädakse vaikimisi nullhüpoteesi juurde ehk loetakse mõõtevahend nõuetele mittevastavaks.

3.1 Klassikaline statistiline käsitus

Kahe ühepoolse teststatistiku leidmist nimetatakse lühidalt *TOST*-testiks (ingl *Two One-Sided Tests*), mille eesmärk on hinnata, kas uuritav efekt jääb etteantud tolerantsipiiride vahele. Erinevalt klassikalisest *t*-testist, kus nullhüpoteesiks on erinevuse puudumine, püstitatakse *TOST*-testis hüpoteesid nii, et nullhüpotees väidab sisuliselt mitteekvivalentsust. Olgu μ_A ja μ_B kahe rühma (näiteks etaloni ja testvihi mõõtmiste) keskväärtused ning nende vaheline erinevus $\Delta M = \mu_A - \mu_B$. Kui ΔM jääb väljapoole tolerantsipiirkonda $(\delta_L; \delta_U)$, kus δ_L tähistab alumist ja δ_U ülemist tolerantsipiiri, loetakse rühmad erinevaks. Testitakse korraga kahte ühepoolset hüpoteesipaari:

$$H_{01} : \mu_A - \mu_B < \delta_L \quad H_{11} : \mu_A - \mu_B \geq \delta_L$$

ja

$$H_{02} : \mu_A - \mu_B > \delta_U \quad H_{12} : \mu_A - \mu_B \leq \delta_U.$$

Rühmad loetakse ekvivalentseks ehk samaväärseks juhul, kui mõlemad ühepoolsed nullhüpoteesid õnnestub valitud olulisuse nivool α tagasi lükata. (Yen, Leber ja Pibida, 2020)

Hüpoteeside testimiseks kasutatakse kahte ühepoolset *t*-testi, mille teststatistikud on

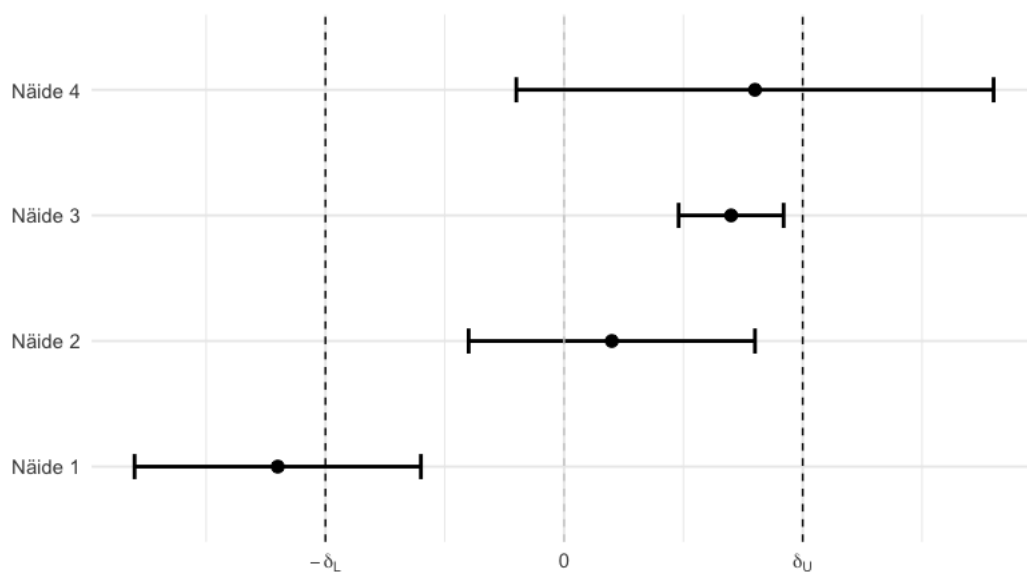
$$t_L = \frac{\widehat{\Delta M} - \delta_L}{\frac{s}{\sqrt{n}}} \quad \text{ja} \quad t_U = \frac{\widehat{\Delta M} - \delta_U}{\frac{s}{\sqrt{n}}}, \quad (5)$$

kus $\widehat{\Delta M} = \hat{\mu}_A - \hat{\mu}_B$ on rühmade keskmiste erinevus, $(\delta_L; \delta_U)$ tolerantsipiirkond, s suuruse $\widehat{\Delta M}$ standardhälve ning n sõltumatute $\widehat{\Delta M}$ mõõtmiste arv. I liiki vea määr kirjeldab ülempiiri tõenäosusele eksklikult lugeda mittesobiv viht nõuetele vastavaks. *TOST*-test on üles ehitatud nii, et selle vea tõenäosus ei ületa valitud olulisuse nivood α . Maksimaalne I tüüpi vea tõenäosus saavutatakse olukorras,

kus tegelik efekt asub täpselt tolerantsipiiril. (Lakens, 2017)

Ekvivalentsust saab analoogselt kontrollida ka usaldusvahemike kaudu. *TOST*-testi puhul kasutatakse 95%-lise usaldusvahemiku (UI, $1 - \alpha$) asemel 90%-list usaldusvahemikku (UI, $1 - 2\alpha$), kuna teostatakse kaks ühepoolset testi (mõlema puhul $\alpha = 0,05$). (Lakens, 2017)

Kui testitakse korraga nii nullhüpoteesi olulisust ($H_0 : \mu_A = \mu_B$) kui ka ekvivalentsust, mille hüpoteesid on välja toodud peatüki alguses, on uuringul neli võimalikku tulemust. Joonisel 3 on nelja stsenaariumi puhul illustreeritud keskmised erinevused ning nende 90%-lised usaldusvahemikud. Must katkendlik joon tähistab tolerantsipiirkonda ($\delta_L; \delta_U$).



Joonis 3: Usaldusintervallide interpreteerimine vastavushindamisel.

Esimese näite puhul on gruppide vahel statistiliselt oluline erinevus, ekvivalentsus puudub. Teise näite puhul suudetakse tõestada samasus ning kuna usaldusvahemik sisaldab nulli, puudub oluline erinevus. Kolmanda näite puhul suudetakse kummutada mõlemad nullhüpoteesid, mis tähendab, et erinevus on statistiliselt

oluline, kuid kuna usaldusintervall asub täielikult tolerantsipiiride vahel, on see erinevus praktilises mõttes tähtsusetu. Viimase stsenaariumi puhul ei ole gruppide vaheline erinevus statistiliselt oluline, ei saa tõestada ka ekvivalentsust.

Statistikatarkvaras R on testi hindamiseks paketi *TOSTER* funktsioon *t_TOST()*, mis töötab sarnaselt R-i *t*-testile, aga viib läbi kolm *t*-testi üheaegselt: ühe kahepoolse ja kaks ühepoolset testi ekvivalentsuse hindamiseks. Funktsioon võtab argumendiks mõõtmistulemused kahest võrreldavast populatsioonist, näiteks kahe erineva kaaluvihi mõõtmiste tulemused, ja tolerantsipiirid ning väljastab *p*-väärtused ja usaldusvahemikud vastavate hüpoteeside jaoks.

3.2 Bayesi käsitus vastavushindamisel

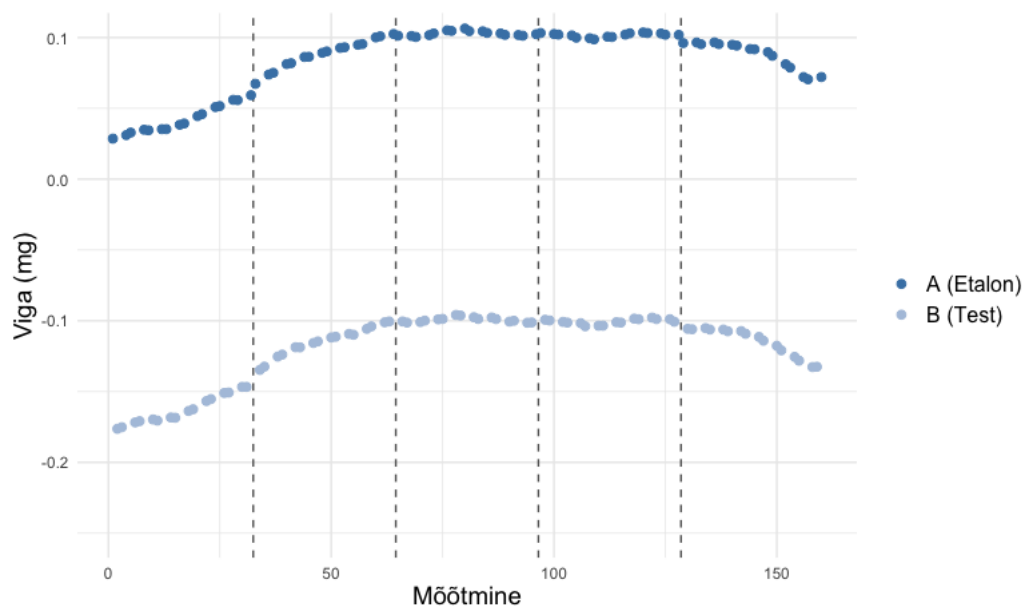
Bayesi peatükis käsitleti ühepoolset hüpoteeside testimist. Ekvivalentsuse testimise korral rakendub sama põhimõte, kuid nullhüpotees koosneb nüüd kahest piirkonnast, mis jäävad väljapoole etteantud tolerantsipiire. Seetõttu leitakse järeldaotuse põhjal tõenäosus, et parameeter jääb alumisest piirist väiksemaks ning tõenäosus, et parameeter ületab ülemise piirväärtuse. Nende tõenäosuste summa annab järeldaotuse tõenäosuse, et nullhüpotees kehtib. (Bolstad ja Curran, 2017)

4 Andmeanalüüs

Järgnevalt teostakse näidisanalüüs kasutades Eesti metroloogia keskasutuse Metrosert andmestikku. Käesoleva töö praktiline osa ei kujuta endast täielikku metroloogilist määramatuse analüüsi, sest ei koostatud määramatuse eelarvet ega modelleeritud eraldi kõiki mõõteprotsessi mõjutavaid määramatuse allikaid, nagu etaloni määramatus, keskkonnategurid või võimalik süsteemne drift. Selline lihtsus võimaldas keskenduda statistiliste otsustusmeetodite võrdlemisele, kuid piirab tulemuste üldistatavust reaalsele kalibreerimispraktikale.

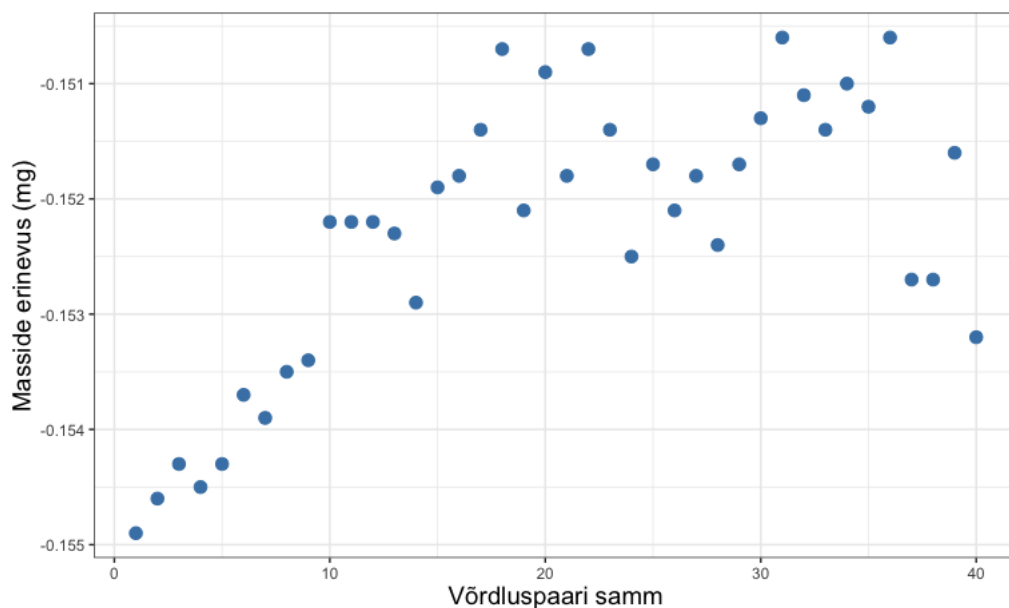
4.1 Andmestiku ülevaade

Kasutatavad andmed on saadud etaloni ja testvihi võrdlusmõõtmise tulemusena. Mõõteprotsessi eesmärk oli määrata 500-grammise testvihi tegelik mass, võrreldes seda etaloniga. Eelnevalt on teada, et etalon on nimiväärtusest hinnanguliselt 0,0507 mg raskem. Käesolevas töös käsitleti etaloni parandust 0,0507 mg lihtsustatult täpse konstandina ning selle määramatust eraldi mudelisse ei kaasatud. Mõõtmised teostati A–B–B–A skeemi kohaselt, kus A tähistab etaloni (referents) kaalumõõtmist ja B testitavat vihti kaalumõõtmist. Ühel päeval tehti kokku 160 üksikmõõtmist, mis vastab 40-le A–B–B–A tsüklile.



Joonis 4: Etaloni ja testvihi kaalude erinevus nominaalväärtusest. Ühe päeva jook-sul tehtud 160 üksikmõõtmise tulemused.

Joonisel 4 on kujutatud vihtide masside hälbed nimiväärtusest. Jooniselt on näha, et andmestik koosneb viiest A–B–B–A tsüklist, millest igaühes on kaheksa mõõtmisseeriat. Mõõtevead järgivad sammudega muutuvat trendi, mis avaldub sarnaselt nii etalonil kui ka testvihil. Võib eeldada, et tulemusi on mõjutanud kõrvalised tegurid, näiteks temperatuur või õhuniiskus. Etaloni kaaluvea hinnangud jäid vahemikku 0,0286 kuni 0,1064 mg ning testvihi vead vahemikku –0,1763 kuni –0,0959 mg.



Joonis 5: Etaloni ja testvihi kaalude erinevus. Ühe mõõtettsükli jooksul saadud etaloni kaalumõõtmiste keskmise ja testvihi kaalumõõtmiste keskmise vahe.

Joonisel 5 on esitatud etaloni ja testvihi vahelised arvutatud vigade erinevused $\widehat{E}[B] - \widehat{E}[A] + 0,0507$ ehk andmepunktid on leitud iga A–B–B–A tsükli lõpus saadud A ja B keskmiste vahena, muuhulgas on arvesse võetud ka etaloni teadaolev viga (0,0507 mg). Selle saab lühemalt kirja panna valemi (1) abil

$$Y = X_1 + X_2,$$

kus X_1 on mõõdetud keskmine erinevus etaloni ja testvihi vahel ($\widehat{\Delta M}$) ja X_2 etaloni teadaolev parandus 0,0507 mg. Seega kujutab Y endast parandatud hinnangut testvihi ja etaloni massi tegelikule erinevusele. Jooniselt 5 on näha, et väärtused jäid vahemikku $-0,1549$ kuni $-0,1507$ mg, kusjuures keskmine erinevus on $-0,1523$ mg. Kuna kõik punktid on negatiivsed, kinnitab see joonisel 4 nähtud erinevust, et testviht on etalonist kergem.

4.2 TOST-test ekvivalentsuse hindamisel

Ekvivalentsuse testimiseks kasutatakse valemeid (5) ning arvutatakse vastavad teststatistikud. Aluseks võeti tabel *OIML R 111-1*, mis lubab 500-grammise E2-klassi vihi maksimaalseks veaks 0,8 mg (International Organization of Legal Metrology, 2004). Seega on tolerantsipiirkonnaks $(-0,8; 0,8)$. Vastavad hüpoteesid, mida kontrollitakse, on

- nullhüpoteesi test

$$H_{01} : \mu_Y = 0 \quad \text{ja} \quad H_{11} : \mu_Y \neq 0;$$

- ekvivalentsustest

$$H_{02} : \mu_Y \leq -0,8 \quad \text{või} \quad \mu_Y \geq 0,8$$

$$H_{12} : -0,8 < \mu_Y < 0,8.$$

Iga A–B–B–A tsükli kohta arvutati erinevus

$$\widehat{\Delta M_i} = \frac{B_{i1} + B_{i2}}{2} - \frac{A_{i1} + A_{i2}}{2},$$

kus B_{ij} näitab i . tsükli testitava vihi j . mõõtmist ja A_{ij} näitab i . tsükli etaloni ehk võrdlusvihi j . mõõtmist. Selle tulemusena saadi $n = 40$ sõltumatut hinnangut. Hinnang kaaluvihide erinevusele i . tsükli mõõtmiste põhjal on

$$Y_i = \widehat{\Delta M_i} + 0,0507,$$

mille põhjal arvutati valimi keskmine ja standardhälve

$$\bar{Y} = \sum_{i=1}^n \frac{Y_i}{n} = -0,1523, \quad s = 0,0012.$$

Seega saadi teststatistikud

$$t_L = \frac{\bar{Y} + 0,8}{\frac{s}{\sqrt{n}}} = 3413,679$$

ja

$$t_U = \frac{\bar{Y} - 0,8}{\frac{s}{\sqrt{n}}} = -5019,062.$$

Kuna $t_{0,95;39} = 1,68$ ning $t_L > t_{0,95;39}$ ja $t_U < -t_{0,95;39}$, siis kummutati mõlemad ühepoolsed nullhüpoteesid.

Kahepoolse t -testi puhul arvutati statistiku väärtuseks

$$t = \frac{\bar{Y} - 0}{\frac{s}{\sqrt{n}}} = -802,691.$$

Seega $|t| = 802,691 > t_{0,975;39} = 2,02$ ja kummutati nullhüpotees (H_{01}).

Kontrolliks viidi sama test läbi ka rakendustarkvaras R, kasutades funktsiooni $t_TOST()$. Analüüsist saadi, et testvihi ja etaloni masside vahe on statistiliselt oluline ($p < 0,001$), kuna p -väärtus on väiksem kui olulisuse nivoo ($\alpha = 0,05$). Hoolimata leitud statistilisest erinevusest, osutus oluliseks ka ekvivalentsustest ($p < 0,001$). Kuna saadud 90%-line usaldusvahemik tuli $(-0,1526; -0,1520) \subset (-0,8; 0,8)$, siis jääb leitud erinevus lubatud piiridesse ehk võib vihid samaväärseks lugeda.

4.3 Bayesi analüüs

Lisaks ekvivalentsustestile rakendati andmetele Bayesi lähenemisviisi, kus testitava vihi massi θ käsitleti juhusliku suurusena. Eeljaotuseks valiti normaaljaotus

$$\theta \sim \mathcal{N}(\mu_{prior}, \sigma_{prior}^2),$$

kus μ_{prior} on eeljaotuse keskväärtus, σ_{prior}^2 eeljaotuse dispersioon ja üksikvaatluseks on üks A–B–B–A mõõtettsükkel. Lisaks eeldati, et enne mõõtmisi asub tegelik massiviga 90%-lise tõenäosusega tolerantipsiirides $(-0,8; 0,8)$. Keskväärtuseks võeti $\mu_{prior} = 0$ mg, kuna kalibreeritud massietaloni ja nõuetele vastava testvihi puhul eeldatakse enne mõõtmisi, et tegelik massiviga on nominaalväärtuse suhtes võimalikult väike. Eeljaotuse standardhälve saadi tingimusest

$$P(-0,8 \leq \theta \leq 0,8) = 0,90$$

ehk

$$\sigma_{prior} = \frac{\delta}{z_{0,95}} = \frac{0,8}{1,645} \approx 0,486 \text{ mg}.$$

Eeldati, et iga mõõtetulemus on normaaljaotusega juhuslik suurus keskväärtusega θ ja dispersiooniga σ^2

$$Y_i \mid \theta \sim \mathcal{N}(\theta, \sigma^2),$$

kus Y_i tähistab parandatud mõõtetulemust ja σ^2 mõõtmiste dispersiooni. Kogu valimi tõepärafunktsioon avaldub kujul

$$L(\theta) \propto \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2 \right).$$

Kasutades seost

$$\sum_{i=1}^n (y_i - \theta)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \theta)^2,$$

jääb θ -st sõltuvaks osaks

$$L(\theta) \propto \exp \left(-\frac{n(\bar{y} - \theta)^2}{2\sigma^2} \right).$$

Kuna eeljaotus on normaaljaotusega, on seda ka järeljaotus parameetritega

$$\theta \mid y \sim \mathcal{N}(\mu_{post}, \sigma_{post}^2),$$

kus

$$\mu_{post} = \frac{\mu_{prior} \frac{\sigma^2}{n} + \bar{y} \sigma_{prior}^2}{\frac{\sigma^2}{n} + \sigma_{prior}^2} \quad \text{ja} \quad \sigma_{post}^2 = \frac{\frac{\sigma^2}{n} \sigma_{prior}^2}{\frac{\sigma^2}{n} + \sigma_{prior}^2}.$$

Eelnevalt on arvatatud $\bar{y} = -0,1523$ mg ja $\sigma = 0,0012$ mg. Seega tuleb järeljaotuse keskväärtuseks

$$\mu_{post} = \frac{0 + (-0,1523) \cdot 0,486^2}{\frac{0,0012^2}{40} + 0,486^2} \approx -0,1523 \text{ mg}$$

ja dispersiooniks

$$\sigma_{post}^2 = \frac{\frac{0,0012^2}{40} \cdot 0,486^2}{\frac{0,0012^2}{40} + 0,486^2} \approx 3,6 \cdot 10^{-8} \text{ mg}^2.$$

Tõenäosusvahemik leitakse normaaljaotuse kvantiili abil

$$\mu_{post} \pm z_{0,975} \sigma_{post},$$

kus $z_{0,975} = 1,96$ ja $\sigma_{post} \approx 0,0002$ mg. Seega on Bayesi 95%-line tõenäosusvahemik $(-0,1527; -0,1519)$.

Tõenäosuse, et tegeliku testvihi mass jääb piiridesse, saab arvutada vastavalt tingimusele

$$P(-0,8 \leq \theta \leq 0,8) = \Phi\left(\frac{0,8 - \mu_{post}}{\sigma_{post}}\right) - \Phi\left(\frac{-0,8 - \mu_{post}}{\sigma_{post}}\right),$$

kus Φ tähistab standardse normaaljaotuse jaotusfunktsiooni. Tõenäosus avaldub kujul

$$P(-0,8 \leq \theta \leq 0,8) = \Phi\left(\frac{0,8 - (-0,1523)}{0,0002}\right) - \Phi\left(\frac{-0,8 - (-0,1523)}{0,0002}\right) \approx 1.$$

See tähendab, et järeldaotuse põhjal jääb tegelik testvihi mass väga suure tõenäosusega etteantud piiridesse.

Täiendavalt vaadeldi analüüsis ka nõrgemat eeljaotust, mille korral jääb tegelik massiviga 50%-lise tõenäosusega tolerantsipiiridest välja, kus eeljaotuse keskvärtuseks valiti uuesti $\mu_{prior} = 0$ mg ja standardhälve saadi tingimusest

$$P(-0,8 \leq \theta \leq 0,8) = 0,50$$

ehk

$$\sigma_{prior} = \frac{\delta}{z_{0,75}} = \frac{0,8}{0,674} \approx 1,187 \text{ mg.}$$

Vastavad järeljaotuse parameetrid tulid

$$\mu_{post} \approx -0,1523 \text{ mg} \quad \text{ja} \quad \sigma_{post}^2 \approx 3,6 \cdot 10^{-8} \text{ mg}^2$$

ning 95%-line tõenäosusvahemik $(-0,1527; -0,1519)$. Kuigi nõrgema eeljaotuse korral suurenes eeljaotuse standardhälve, jäid järeljaotuse parameetrid praktiliselt samaks.

4.4 A–B–B–A mõõtettsükli simulatsioon

Meetodite võrdlemiseks viidi läbi simulatsioon, milles modelleeriti kõigepealt testvihi tegelik massiviga θ_{true} , mille alusel otsustati, kas viht on tegelikult nõue-tele vastav:

$$\begin{cases} 1, & \text{kui } |\theta_{true}| \leq 0,8, \\ 0, & \text{kui } |\theta_{true}| > 0,8. \end{cases}$$

Seejärel genereeriti A–B–B–A üksikmõõtmised. Kui etaloni teadaolev viga on tähistatud sümboliga $\gamma = 0,0507$ mg ning üksikmõõtmiste juhuslik mõõteviga on normaaljaotusega

$$\eta \sim \mathcal{N}(0, \sigma_w^2),$$

siis ühe tsükli neli mõõtmist on

$$A_1 = \gamma + \eta_{A1},$$

$$B_1 = \theta_{true} + \eta_{B1},$$

$$B_2 = \theta_{true} + \eta_{B2},$$

$$A_2 = \gamma + \eta_{A2}.$$

Tsükli lõpus arvutati etaloni ja testvihi keskmised ning nende vahe

$$\widehat{\Delta M} = \frac{B_1 + B_2}{2} - \frac{A_1 + A_2}{2} = \bar{B} - \bar{A}.$$

Lõplik parandatud andmepunkt oli

$$Y = \widehat{\Delta M} + \gamma.$$

Asendades $\bar{A}$ ja $\bar{B}$ avaldised, saame

$$Y = \theta_{\text{true}} + \frac{\eta_{B1} + \eta_{B2} - \eta_{A1} - \eta_{A2}}{2}.$$

Kokku genereeriti sama θ_{true} väärtust kasutades 40 mõõtmist Y . Nende mõõtmistulemuste põhjal hinnati parameetrit θ_{true} nii klassikalise kui ka Bayesi meetodi abil.

4.4.1 Simulatsiooni ülesehitus

Simulatsioonis korraldati samme 1–4 1000 korda:

1. Genereeriti testvihi tegelik massiviga θ_{true} eelnevalt määratud jaotusest ja otsustati, kas viht oli tegelikult lubatud vahemikus või mitte.
2. Genereeriti A–B–B–A skeemi järgi mõõtmistulemused, kokku kaheksa seeriat, kus igas seerias oli viis tsükli.
3. Arvutati simuleeritud andmete põhjal testvihi massivea hinnang ning rakendati *TOST*-test, et otsustada, kas testviht vastab nõuetele.
4. Sama simuleeritud andmestiku põhjal arvutati Bayesi järeljaotus ja tehti otsus nõuetele vastavuse kohta.

4.4.2 Hindamiskriteeriumid

TOST-test ja Bayesi lähenemine lahendavad mõlemad sama praktilist ülesannet, kuid teevad seda erineva raamistiku alusel. *TOST*-test võimaldab kontrollida, kas vaatluste põhjal saadud hinnang ja selle määramatus on kooskõlas nõudega, et tegelik massiviga jääb vahemikku $(\delta_L; \delta_U)$. Bayesi meetod võimaldab seevastu väljendada sama küsimust järeljaotuse tõenäosusena, st kui tõenäoline on pärast mõõtmisi ja eelinfo arvestamist, et testviht on tegelikult lubatud piirides. Eeljaot-

tus valitakse samal eeldusel nagu eelnevas näites, et tegelik viga asub 90%-lise tõenäosusega tolerantsipiirides $(-0,8; 0,8)$.

Hinnangute täpsuse kirjeldamiseks kasutati keskmist ruutviga (ingl *mean squared error*)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \left(\hat{\theta}_i - \theta_{\text{true},i} \right)^2,$$

kus $\hat{\theta}_i$ tähistab i -ndat hinnangut ning N simulatsioonikorduste arvu. Mida väiksem on MSE, seda täpsem on meetod.

Lisaks hinnati meetodite otsustuskvaliteeti tundlikkuse ja spetsiifilisuse abil. Tundlikkus kirjeldab, kui sageli tuvastab meetod õigesti nõuetele vastava vihi. Empiiriliselt hinnati seda valemiga

$$\widehat{\text{Tundlikkus}} = \frac{\#\{\text{õigesti vastavaks loetud vihid}\}}{\#\{\text{tegelikult vastavad vihid}\}}. \quad (6)$$

Spetsiifilisus kirjeldab, kui sageli tuvastab meetod nõuetele mittevastava vihi õigesti mittevastavaks. Empiiriliselt hinnati seda valemiga

$$\widehat{\text{Spetsiifilisus}} = \frac{\#\{\text{õigesti mittevastavaks loetud vihid}\}}{\#\{\text{tegelikult mittevastavad vihid}\}}. \quad (7)$$

Täiendavalt võib vajaduse korral vaadelda ka valepositiivsete ja valenegatiivsete otsuste sagedust. Vastavushindamise kontekstis tähendab valepositiivne otsus olukorda, kus nõuetele mittevastav viht loetakse ekslikult vastavaks, ning valenegatiivne otsus olukorda, kus tegelikult nõuetele vastav viht loetakse ekslikult mittevastavaks. (JCGM, 2012)

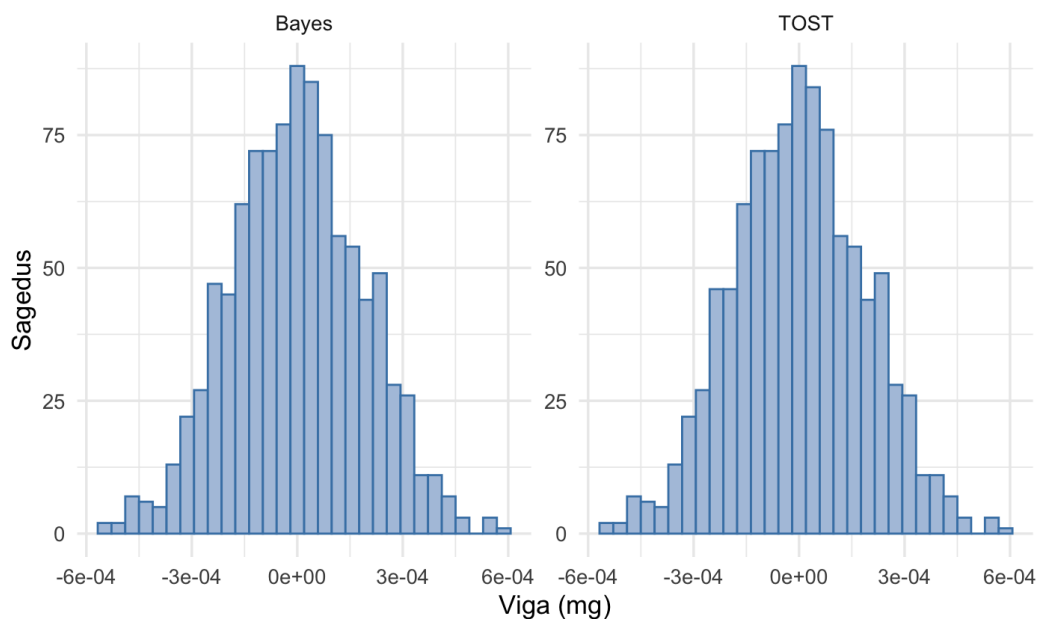
4.4.3 Simulatsiooni tulemused

Simulatsioonis genereeriti kokku 1000 vihti nii, et 90% (891) olid nõuetele vastavad ja ülejäänud 10% (109) mittevastavad. Mõlemad meetodid klassifitseerisid 891 vihti vastavaks ja 109 mittevastavaks. Seda kinnitavad ka arvutatud kvaliteedinäitajad, mis on välja toodud allolevas tabelis.

Tabel 2: *TOST*- ja Bayesi meetodi simulatsiooni tulemused.

Meetod	MSE	Tundlikkus	Spetsiifilisus
TOST	$3,571792 \cdot 10^{-8}$	1	1
Bayes	$3,571771 \cdot 10^{-8}$	1	1

Keskmine ruutviga (MSE) tuli mõlemal meetodil praktiliselt sama ning väga väike, mis näitab, et hinnangud θ_{true} -le olid väga täpsed. Tundlikkus ja spetsiifilisus tulid mõlemad 1, mis tähendab, et kõik vihid klassifitseeriti korrektselt.



Joonis 6: Parameetri hinnangu vigade jaotused mõlema meetodi puhul.

Nagu jooniselt 6 näha, on mõlema meetodi vead praktiliselt sama jaotusega. Seega

võib järeldada, et antud tingimustes ei anna Bayesi lähenemine võrreldes *TOST*-testiga sisulist eelist. Tuleb siiski märkida, et käesolevas töös jäi eeljaotuse mõju järeljaotusele väga väikeseks, kuna mõõtmiste arv oli suhteliselt suur ning mõõtemüra väike võrreldes etteantud tolerantspiiridega. Sellisel juhul domineerib Bayesi analüüsis tõepärafunktsioon ning järeljaotus kujuneb peamiselt vaatlusandmete põhjal. Väiksemate valimite, suurema mõõtemääramatuse või tugevama eelinfo korral võib eeljaotuse valik mõjutada tulemusi oluliselt rohkem. Seetõttu on Bayesi meetodite puhul oluline hinnata ka seda, kui tundlikud on saadud järeldused eeljaotuse valiku suhtes.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärk oli võrrelda klassikalist ja Bayesi lähenemist vastavushindamise kontekstis. Lisaks anti ülevaade mõõtemääramatuse hindamise põhimõtetest ning vastavushindamisel kasutatavatest statistilistest meetoditest.

Töö teoreetilises osas käsitleti sageduslikku ja Bayesi statistikat, analüüsides nende peamisi erinevusi tõenäosuse tõlgendamisel ning statistiliste järelduste tegemisel. Selgitati, kuidas Bayesi lähenemine võimaldab lisaks vaatlusandmetele kaasata eelteavet ning kirjeldada mõõdetavat parameetrit järeljaotuse kaudu. Metroloogilisest aspektist käsitleti mõõtemääramatust ja vastavushindamist.

Praktilises osas rakendati *TOST*-testi ja Bayesi meetodeid Metroserdi näidisandmestikul. Analüüsist selgus, et kuigi testvihi ja etaloni vahel esines statistiliselt oluline erinevus, jäi see lubatud tolerantsipiiridesse, mis kinnitab vihi vastavust nõuetele. Bayesi analüüs kinnitas antud järeldust, osutades väga suurele tõenäosusele, et massi tegelik väärtus jääb lubatud vahemikku.

Lisaks uuriti simulatsiooni abil A–B–B–A mõõtettsükli korral *TOST*- ja Bayesi meetodi käitumist. Simulatsiooni tulemused näitasid, et vaadeldud tingimustes andsid mõlemad meetodid praktiliselt samad tulemused. Mõlema meetodi hinnangud olid täpsed ning kõik simuleeritud vihid klassifitseeriti korrektselt. Kuna mõõtemüra oli võrreldes tolerantsiga väike ja mõõtmiste arv suur, määras tulemu-
se peamiselt tõepärafunktsioon ning eeljaotuse mõju jäi väikseks.

Kasutatud allikad

- Bolstad, William M. ja James M. Curran (2017). *Introduction to Bayesian Statistics, Third Edition*. Wiley.
- International Organization of Legal Metrology (2004). *OIML R 111-1: Weights of classes E1, E2, F1, F2, M1, M1-2, M2, M2-3 and M3 – Part 1: Metrological and technical requirements*. URL: https://www.oiml.org/en/files/pdf_r/r111-1-e04.pdf.
- JCGM (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM)*. BIPM. URL: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf.
- JCGM (2012). *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment*. BIPM. URL: https://www.bipm.org/documents/20126/2071204/JCGM_106_2012_E.pdf.
- Lakens, Daniël (2017). „Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses“. *Social Psychological and Personality Science* 8.4, lk. 355–362. DOI: [10.1177/1948550617697177](https://doi.org/10.1177/1948550617697177).
- Possolo, A. (2015). *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. National Institute of Standards ja Technology. DOI: [10.6028/NIST.TN.1900](https://doi.org/10.6028/NIST.TN.1900).
- Wei, Zhengxiao, Aijun Yang, Leno Rocha, Michelle F. Miranda ja Farouk S. Nathoo (2022). „A Review of Bayesian Hypothesis Testing and Its Practical Implementations“. *Entropy* 24.2, lk. 161. DOI: [10.3390/e24020161](https://doi.org/10.3390/e24020161).
- Willink, R. ja R. White (s.a.). *Disentangling Classical and Bayesian Approaches to Uncertainty Analysis*. URL: <https://www.bipm.org/documents/20126/28435677/working-document-ID-5058/8fd49ede-0e53-66b1-050a-edae2c2f62c>.

Yen, J., D. Leber ja L. Pibida (2020). *Comparing Instruments*. National Institute of Standards ja Technology. DOI: [10.6028/NIST.TN.2106](https://doi.org/10.6028/NIST.TN.2106).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Mari-Ann Liidemaa,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Vastavushindamine metroloogias: Bayesi ja klassikalise statistika võrdlev käsitus“, mille juhendajad on Aigar Vaigu ja Märt Möls, 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.

Mari-Ann Liidemaa

12.05.2026