

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MATEMAATIKA JA STATISTIKA INSTITUUT

Mihhail Bobeško  
**Astakute ja pseudo-astakute võrdlus**  
**Kruskal-Wallise testi korral**

Matemaatiline statistika  
Bakalaureusetöö (9 EAP)

Juhendaja: MSc Anne Selart

TARTU 2025

**ASTAKUTE JA PSEUDO-ASTAKUTE VÕRDLUS  
KRUSKAL-WALLISE TESTI KORRAL**

Bakalaureusetöö  
Mihhail Bobeško

**Lühikokkuvõte**

Bakalaureusetöös uuritakse pseudo-astakute kasutamist mitteparameetrites testides ning võrreldakse neid traditsiooniliste astakutel põhineva testiga. Töö keskendub olukordadele, kus rühmade valimimahud on ebavõrdsed ja klassikalised testid (nt Kruskal–Wallise test) võivad anda eksitavaid tulemusi.

Teoreetilises osas tutvustatakse astakute ja pseudo-astakute mõistet, suhtelisi efekte ning nende arvutuslikku tausta. Praktilises osas viiakse läbi simulatsioonid normaaljaotuste segudega, et hinnata testide tundlikkust valimimahu muutuste suhtes ning illustreerida mittetransitiivseid olukordi.

Töö tulemusena selgub, et pseudo-astakutel põhinev test on usaldusväärsem suurte ebavõrdsete valimite korral, kuna see ei sõltu rühmade suurusest.

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

**Märksõnad:** mitteparameetiline statistika, pseudo-astakud, Kruskal-Wallise test, simulatsioonid.

**COMPARISON OF RANKS AND PSEUDO-RANKS IN THE  
KRUSKAL–WALLIS TEST**

Bachelor thesis  
Mihhail Bobeško

**Abstract**

This bachelor's thesis investigates the use of pseudo-ranks in nonparametric

test and compares them to classical rank-based method. The focus is on situations where group sample sizes are unequal, which may lead traditional tests (such as the Kruskal–Wallis test) to produce misleading results.

The theoretical part introduces the concepts of ranks and pseudo-ranks, relative effects, and their computation. In the practical part, simulations with mixtures of normal distributions are conducted to assess the sensitivity of test results to imbalanced sample sizes and to demonstrate nontransitive scenarios.

The results show that tests based on pseudo-ranks provide more reliable conclusions in the presence of unequal group sizes, as they do not depend on the sample size proportions.

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

**Key Words:** nonparametric statistics, pseudo-ranks, Kruskal–Wallis test, simulations.

# Sisukord

<b>Sissejuhatus</b>	<b>5</b>
<b>1 Astakud ja pseudo-astakud: definitsioon, tähistused ja kasutamine</b>	<b>6</b>
1.1 Mitteparameetrilised meetodid statistikas . . . . .	6
1.2 Pseudo-astak . . . . .	6
1.3 Mitme rühma võrdlemise katse korraldus ja tähistused . . . . .	7
1.4 Kaalutud suhtelised efektid ja astakud . . . . .	7
1.5 Kaalumata suhtelised efektid ja pseudo-astakud . . . . .	14
<b>2 Astakute ja pseudo-astakute arvutamise näide</b>	<b>18</b>
2.1 Funktsiooni $c(u)$ väärtused . . . . .	18
2.2 Pseudo-astakute arvutamine . . . . .	19
<b>3 Astakutel põhinevate testide ootamatud tulemused</b>	<b>22</b>
3.1 Mittetransiivne struktuur segujaotuste korral . . . . .	22
3.2 Simulatsioonid: klassikalise ja pseudo-astakutel põhineva Kruskal–Wallise testi võrdlus . . . . .	23
<b>Kokkuvõte</b>	<b>29</b>
<b>Kasutatud allikad</b>	<b>30</b>
<b>Lisa 1. Kaalutud suhteliste efektide arvutamine</b>	<b>31</b>
<b>Lisa 2. Pseudo-astakute arvutamine tarkvara R abil</b>	<b>35</b>
<b>Lisa 3. Testide ootamatud tulemused: pidevate jaotustega näide</b>	<b>36</b>

**Lisa 4. Klassikalise Kruskal-Wallise testi ja pseudo-astakutel põhineva  
Kruskal-Wallise testi võrdlemine**

**41**

## Sissejuhatus

Statistikas kasutatakse andmete analüüsimisel nii parameetrilisi kui ka mitteparameetrilisi meetodeid. Parameetrilised testid eeldavad sageli, et andmed pärinevad teatud tüüpi jaotusest (näiteks normaaljaotusest) ning et rühmade hajuvused on sarnased. Kui need eeldused ei ole täidetud või andmete jaotused ei ole teada, on sobivamad mitteparameetrilised meetodid, mis ei sõltu jaotuse kujust ning põhinevad sageli andmete järjestusel ehk astakutel.

Üks levinumaid mitteparameetrilisi meetodeid on Kruskal–Wallise test, mida kasutatakse mitme sõltumatu rühma võrdlemiseks. Testi tugevuseks on selle lihtsus ja vähene tundlikkus jaotuse kuju suhtes, kuid probleemiks võib osutuda selle tundlikkus valimimahu ebavõrdsuse suhtes. Kui rühmade suurused on väga erinevad, võib see mõjutada teststatistiku väärtust ning viia eksitavate järeldusteni.

Selliste olukordade jaoks on välja töötatud pseudo-astakutel põhinev lähenemine, mis võimaldab võrrelda rühmi usaldusväärsemalt. Pseudo-astakud kirjeldavad rühmade suhtelisi efekte viisil, mis ei sõltu otseselt valimimahust. Seetõttu on pseudo-astakutel põhinev test stabiilsem ja sobivam kasutamiseks just siis, kui andmed on tasakaalustamata.

Käesoleva töö eesmärk on uurida pseudo-astakute omadusi ning nende kasutamist mitme rühma võrdlemisel, tuues esile nende eelised võrreldes klassikaliste astakutel põhineva Kruskal-Wallise testiga.

Käesolevas töös esitatud joonised ja tulemused on saadud tarkvara R abil (R Core Team, 2024). Koodide kirjutamiseks abiks kasutati tehisintellekti (OpenAI ChatGPT, 2025). Bakalaureusetöö on vormistatud L<sup>A</sup>T<sub>E</sub>Xiga, kasutades Overleaf rakendust.

# 1 Astakud ja pseudo-astakud: definitsioon, tähistused ja kasutamine

Järgnev peatükk on kirjutatud artikli (Brunner *et al.*, 2020) põhjal, kui ei ole märgitud teisiti.

Selles peatükis tutvustatakse astaku ja pseudo-astaku mõistet, vajalikke tähistusi ja kasutamise võimalusi.

## 1.1 Mitteparameetrilised meetodid statistikas

Kui klassikaliste parameetriliste meetodite rakendamiseks vajalikud eeldused ei ole täidetud, peetakse sobivamaks kasutada mitteparameetrilisi meetodeid, mis põhinevad sageli astakutel. Mitteparameetrilised astakupõhised meetodid nõuavad vähem eeldusi kui parameetrilised meetodid, kuid neil on siiski teatud tingimused, mida tuleks arvestada. Sageli väidetakse, et need meetodid on robustsemad ja pakuvad usaldusväärsemaid tulemusi võrreldes nende parameetriliste alternatiividega.

## 1.2 Pseudo-astak

Pseudo-astak on mitteparameetriline statistiline meetod, mis tuleneb klassikalistest astakutest, kuid millel on mitmeid olulisi eeliseid võrreldes tavapärase astakupõhiste testidega. Peamised pseudo-astakute eelised on järgmised:

- pseudo-astakud on vähem tundlikud valimi suuruse erinevustele;
- pseudo-astakud võimaldavad usaldusväärsemat võrdlust, eriti mitme rühma korral;
- pseudo-astakud annavad täpsema hinnangu kaalumata suhtelistele efektidele.

Erinevalt tavapäraastest astakutest, mis hindavad kaalutud suhtelisi efekte, võimaldavad pseudo-astakud mõõta kaalumata suhtelisi efekte, mis tagavad parema võrreldavuse olukordades, kus rühmade valimimahud on erinevad.

Järgnevalt käsitletakse astakute ja pseudo-astakute täpsemat määratlust, nende matemaatilist mudelit ning praktilisi rakendusi.

### 1.3 Mitme rühma võrdlemise katse korraldus ja tähistused

Olgu  $d \geq 3$  sõltumatut rühma, mida tähistatakse indeksiga  $i = 1, \dots, d$ . Igast rühmast võetud valim on suurusega  $n_i$ , vaatluste koguarv on  $N = \sum_{i=1}^d n_i$ . Vaadeldakse tunnust  $X$ , mille jaotus sõltub sellest, millisesse rühma konkreetne vaatlus kuulub. Olgu  $X_{ik}$   $i$ -nda rühma  $k$ -s vaatlus, kus  $i = 1, \dots, d$  ning  $k = 1, \dots, n_i$ . Eeldatakse, et kõik vaatlused on sõltumatud. Iga rühma  $i$  jaotust kirjeldab jaotusfunktsioon  $F_i$ , mis defineeritakse kui vasak- ja parempoolsete jaotusfunktsioonide keskmine:

$$F_i = \frac{1}{2}[F_i^- + F_i^+].$$

### 1.4 Kaalutud suhtelised efektid ja astakud

Järgnevas käsitletakse astakute ja kaalutud suhteliste efektide rolli mitme sõltumatu rühma võrdlemisel, kus igal rühmal võib olla erinev jaotusfunktsioon.

Mitme rühma võrdlemisel kasutatakse suhtelisi efekte,  $i$ -nda rühma suhteline efekt tähistatakse sümboliga  $p_i$ .

Suhtelisi efekte  $p_i$  nimetatakse **kaalutud suhtelisteks efektideks**, kui nende arvutamisel lähtutakse segujaotusest  $H$ , kus iga rühma panus vastab tema valimimahule.

Kaalutud suhtelise efekti  $p_i$  defineerimiseks võetakse aluseks juhuslikult valitud

vaatlus  $Z$ , mis on sõltumatu vaatlusest  $X_{ik}$  ning mille jaotus on segu kõigi rühmade jaotustest  $F_i$ . **Segujaotust** tähistatakse sümboliga  $H$  ning defineeritakse kui rühmade jaotuste kaalutud keskmine:

$$H = \frac{1}{N} \sum_{i=1}^d n_i F_i.$$

Vaatlus  $Z \sim H$  on esindaja, mis peegeldab kogu andmestiku üldist jaotust, võttes arvesse nii rühmade jaotusi kui ka nende suuruseid valimis.

Kaalutud suhteline efekt  $p_i$  väljendab tõenäosust, et vaatlus  $Z$  on väiksem kui juhuslikult valitud vaatlus rühmast  $i$  (näiteks võib kasutada esimest vaatlust  $X_{i1}$ ) ning lisaks pool tõenäosusest, et need on võrdsed:

$$p_i = P(Z < X_{i1}) + \frac{1}{2}P(Z = X_{i1}).$$

Kasutades jaotusfunktsiooni  $H$  võib selle tõenäosuse kirja panna valemiga

$$p_i = \int H dF_i,$$

kus  $H$  on segujaotus ja  $F_i$  on rühma  $i$  jaotusfunktsioon (Brunner *et al.*, 2018).

### Näide: kaalutud suhteliste efektide arvutamine

Järgnevalt on toodud näide, mis illustreerib kaalutud suhteliste efektide arvutamist.

Olgu antud 3 sõltumatut rühma, mille vaatlused on pärit normaaljaotustest  $\mathcal{N}(\mu_i, \sigma_i^2)$ :

- Rühm 1: keskväärtus  $\mu_1 = 0$ , standardhälve  $\sigma_1 = 1$ , valimi suurus  $n_1 = 30$ ;
- Rühm 2:  $\mu_2 = 2$ ,  $\sigma_2 = 1$ ,  $n_2 = 30$ ;
- Rühm 3:  $\mu_3 = 5$ ,  $\sigma_3 = 1$ ,  $n_3 = 40$ .

Kogu valimi suurus on  $N = 30 + 30 + 40 = 100$ .

Kui võtta kolmes rühmas leitud jaotused ning kaaluda need vastavalt iga rühma valimisuurusele, siis saadakse segujaotus  $H$ . Kuna rühmadel on valimi suurused 30, 30 ja 40, siis on nende kaalud:

$$\omega_1 = \frac{30}{100} = 0,3; \quad \omega_2 = \frac{30}{100} = 0,3; \quad \omega_3 = \frac{40}{100} = 0,4.$$

Segujaotus  $H$  kohal  $x$  on defineeritud valemiga

$$H(x) = \frac{1}{N} \sum_{i=1}^d n_i F_i(x) = \omega_1 F_1(x) + \omega_2 F_2(x) + \omega_3 F_3(x),$$

kus  $F_i$  on rühma  $i$  jaotusfunktsioon.

Suhtelise efekti definitsiooni järgi on teada, et  $i$ -nda rühma suhteline efekt on

$$p_i = P(Z < X_i) + \frac{1}{2}P(Z = X_i),$$

kus  $Z \sim H$  on juhuslik suurus segust,  $X_i \sim F_i$  pärineb rühma  $i$  jaotusest.

Kuna tegemist on pidevate jaotustega, siis võrdsete väärtuste tekkimise tõenäosus on  $P(Z = X_i) = 0$ , mistõttu võib kirjutada  $p_i = P(Z < X_i)$ .

Kuna  $Z$  on juhuslik suurus, mis on pärit segujaotusest  $H$ , st see valitakse juhuslikult kolmest rühmajaotusest  $F_r$  vastavalt kaaludele  $\omega_r$ ,  $r = 1, 2, 3$ , kehtib:

$$p_i = \sum_{r=1}^3 \omega_r P(X_r < X_i) = \omega_1 P(X_1 < X_i) + \omega_2 P(X_2 < X_i) + \omega_3 P(X_3 < X_i),$$

kus  $X_r \sim F_r$ .

Kui  $X_r$  ja  $X_i$  on pärit samast pidevast jaotusest ehk  $r = i$ , siis kehtib  $P(X_i < X_i) = \frac{1}{2}$ .

Kui  $\mu_r \neq \mu_i$  ja  $\sigma_r, \sigma_i$  on teada, leitakse  $P(X_r < X_i)$  lähtudes normaaljaotuste omadustest.

Kui  $X_r \sim \mathcal{N}(\mu_r, \sigma_r^2)$  ja  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ , siis  $X_i - X_r \sim \mathcal{N}(\mu_i - \mu_r; \sigma_i^2 + \sigma_r^2)$  ja

$$P(X_r < X_i) = \Phi\left(\frac{\mu_i - \mu_r}{\sqrt{\sigma_i^2 + \sigma_r^2}}\right),$$

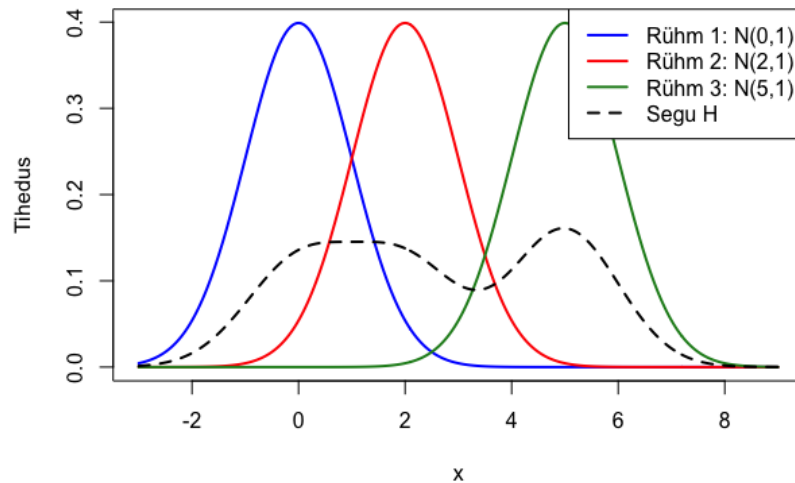
kus  $\Phi(z)$  on standardse normaaljaotuse jaotusfunktsioon.

Põhjalikum arvutuskäik on esitatud lisas [Kaalutud suhteliste efektide arvutamine](#).

Saadud kaalutud suhtelised efektid on:

$$p_1 \approx 0,174; \quad p_2 \approx 0,433; \quad p_3 \approx 0,795.$$

Kaalutud suhtelised efektid kirjeldavad, kui suure tõenäosusega on iga rühma vaatlus väiksem kui juhuslikult valitud vaatlus segujaotusest  $H$ . Näiteks  $p_1 = 0,174$  näitab, et 1. rühma vaatlused paiknevad peamiselt üldise jaotuse alumises osas. Seevastu  $p_3 = 0,795$  tähendab, et 3. rühma vaatlused on üldjuhul suuremad kui ülejäänud segus. Jooniselt (1) on näha iga rühma normaaljaotuse ja segujaotuse tihedusfunktsioon.



Joonis 1: Kolme rühma ja segujaotuse tihedusfunktsioonide graafik

Astakutel põhinev analüüs on sobiv, kui andmete jaotusi ei saa kirjeldada kindla parameetrilise mudeliga. Tavaliselt eeldatakse, et vaatlused on sõltumatud ning et rühmade jaotused võivad üksteisest erineda. **Astakud** tähistavad vaatluste järjekorranumbreid kogu andmestiku ühises järjestuses ehk variatsioonireas. Kasutades segujaotuse  $H$  empiirilist hinnangut  $\hat{H}$ , saab  $i$ -nda rühma  $k$ -nda vaatlusväärtuse  $X_{ik}$  astaku  $R_{ik}$  esitada kujul:

$$R_{ik} = \frac{1}{2} + N\hat{H}(X_{ik}),$$

kus  $N$  on vaatluste koguarv ning empiiriline hinnang segujaotusele  $\hat{H}$  punktis  $x$  on arvatatud kujul:

$$\hat{H}(x) = \sum_{r=1}^d \sum_{\ell=1}^{n_r} c(x - X_{r\ell}),$$

kus funktsioon  $c(u)$  on defineeritud kui

$$c(u) = \begin{cases} 1, & u > 0; \\ 0,5, & u = 0; \\ 0, & u < 0. \end{cases} \quad (1)$$

**Empiiriline jaotusfunktsioon**  $\hat{F}_i(x)$  on mitteparameetriline hinnang  $i$ -nda rühma jaotusfunktsioonile  $F_i(x)$ . See näitab, kui suur osa rühma  $i$  vaatlustest on väiksemad või võrdsed väärtusega  $x$ .

Funktsioon  $\hat{F}_i(x)$  saab esitada funktsiooni (1) abil järgmiselt:

$$\hat{F}_i(x) = \frac{1}{n_i} \sum_{\ell=1}^{n_i} c(x - X_{i\ell}).$$

Seda hinnangut saab rakendada mistahes väärtusele  $x$ , sealhulgas ka konkreetsele vaatlusväärtusele  $X_{ik}$ . Sellisel juhul näitab  $\hat{F}_i(X_{ik})$  osakaalu rühma  $i$  vaatlustest, mis on väiksemad või võrdsed  $X_{ik}$ -ga. Empiiriline jaotusfunktsioon on aluseks

pseudo-astakute ja suhteliste efektide arvutamisel.

**Kaalutud suhtelised efektid** mängivad olulist rolli mitmetes mitteparameetrilistes testides, sealhulgas **Kruskal–Wallise testis**, mis on Wilcoxon astaktesti üldistus mitmele rühmale. See test võimaldab hinnata, kas rühmade jaotused erinevad üksteisest (Brunner *et al.*, 2018).

Kaalutud suhtelise efekti  $p_i$  **hinnang** põhineb empiirilisel segujaotusfunktsioonil  $\hat{H}$  ja rühma  $i$  empiirilisel jaotusfunktsioonil  $\hat{F}_i$ :

$$\hat{p}_i = \int \hat{H} d\hat{F}_i.$$

Alternatiivselt saab suhtelise efekti hinnangu siduda ka astakutega:

$$\hat{p}_i = \frac{1}{N}(\bar{R}_i - \frac{1}{2}), \quad (2)$$

kus  $\bar{R}_i$  on  $i$ -nda rühma vaatluste astakute aritmeetiline keskmine.

**Segujaotus  $H$**  tähendab, et kogu andmestiku jaotusfunktsioon saadakse erinevate rühmade jaotuste  $F_i$  kaalutud summana, kus kaaludeks on rühmade valimi suurused  $n_i$ . Selliselt konstrueeritud segujaotus  $H$  võimaldab arvestada nii jaotuste kuju kui ka rühmade suurusi.

Sellest seosest tulenevalt on võimalik **Kruskal–Wallise teststatistikut  $K$**  väljendada kaalutud suhteliste efektide  $\hat{p}_i$  kaudu. Kuna statistiku valemis esinevad rühmade astakute aritmeetilised keskmised on seotud  $\hat{p}_i$  väärtustega, saab teha järgmise teisenduse.

Kruskal-Wallise teststatistiku klassikaline kuju on

$$K = \frac{12}{N(N+1)} \sum_{i=1}^d n_i \left( \bar{R}_i - \frac{N+1}{2} \right)^2, \quad (3)$$

kus  $d$  on rühmade arv,  $n_i$  on  $i$ -nda rühma suurus,  $\bar{R}_i$  on  $i$ -nda rühma keskmine astak,  $N$  on kõigi rühmade vaatluste koguarv.

Valemi (2) samaväärne esitus on  $\bar{R}_i = N\hat{p}_i + \frac{1}{2}$ . Asendades selle Kruskal-Wallise teststatistiku valemisse (3) saadakse:

$$K = \frac{12}{N(N+1)} \sum_{i=1}^d n_i \left( N\hat{p}_i + \frac{1}{2} - \frac{N+1}{2} \right)^2.$$

Lihtsustades:

$$K = \frac{12N}{(N+1)} \sum_{i=1}^d n_i \left( \hat{p}_i - \frac{1}{2} \right)^2.$$

See teisendus näitab, et Kruskal-Wallise test kontrollib, kui palju kaalutud suhtelised efektid  $\hat{p}_i$  erinevad väärtusest 0,5. Mida suuremad on need erinevused, seda suurem on  $K$  ja seda tõenäolisem on nullhüpoteesi tagasilükkamine.

Kui kõikide rühmade jaotused on samad (ehk kehtib nullhüpotees  $H_0$ ), siis iga rühma kaalutud suhteline efekt on  $p_i = \frac{1}{2}$  ning on oodatud, et valimist arvatud Kruskal-Wallise teststatistiku väärtus on nullilähedane. Siiski võib esineda olukord, kus rühmade jaotused on erinevad (ehk kehtib  $H_1$ ), kuid valimimahud on sellised, et kaalutud suhtelised efektid  $p_i$  jäävad  $\frac{1}{2}$  juurde. Sellisel juhul Kruskal-Wallise teststatistiku väärtused tulevad madalad, mis tähendab, et otsuseks tuleb nullhüpoteesi juurde jääda, kuigi kehtib  $H_1$ . Valimimahtude muutes aga  $p_i \neq \frac{1}{2}$  (kuigi rühmade jaotuste osas midagi ei muutu) ja Kruskal-Wallise teststatistiku väärtused võivad viidata  $H_1$  kehtimisele. Seega astakutel põhineva Kruskal-Wallise teststatistiku puuduseks on see, et selle väärtus sõltub mitte üksnes rühmade jaotustest, vaid ka valimimahtudest. Selle probleemi lahendamiseks on välja pakutud kasutada pseudo-astakuid, mis põhinevad kaalumata suhtelistel efektidel.

Järgmises osas käsitletakse **kaalumata suhtelisi efekte** ja **pseudo-astakuid**, mis võimaldavad hinnata rühmade erinevusi sõltumatult valimimahtudest ning tagavad seetõttu usaldusväärsema tulemuse olukordades, kus rühmade suurused on erinevad.

## 1.5 Kaalumata suhtelised efektid ja pseudo-astakud

Kui eelmises alapunktis kasutati segujaotust  $H$ , mille kaalud põhinesid rühmade valimi suurustel  $n_i$ , siis siinkohal käsitletakse **kaalumata segujaotust**  $G$ , kus iga rühm panustab võrdselt. Seega on mõlemad  $H$  ja  $G$  segujaotused, kuid erinevate kaaludega:  $H$  kasutab empiirilisi proportsioone  $\frac{n_i}{N}$ ,  $G$  aga fikseeritud osakaalu  $\frac{1}{d}$  iga rühma jaoks, kus  $d$  on rühmade koguarv.

**Kaalumata suhtelised efektid** ei sõltu valimimahust ning põhinevad võrdsete kaaludega segujaotusel  $G$ , mis on defineeritud järgmiselt:

$$G = \frac{1}{d} \sum_{i=1}^d F_i,$$

kus  $F_i$  tähistab rühma  $i$  jaotusfunktsiooni.

**Rühma  $i$  kaalumata suhteline efekt**  $\psi_i$  on defineeritud kui

$$\psi_i = \int G dF_i,$$

ehk tõenäosus, et juhuslikult valitud vaatlus  $Z \sim G$  on väiksem kui vaatlus  $X_{i1} \sim F_i$  ning pool tõenäosust, et need on võrdsed:

$$\psi_i = P(Z < X_{i1}) + \frac{1}{2}P(Z = X_{i1}).$$

Kaalumata suhtelise efekti empiiriline hinnang  $\hat{\psi}_i$  arvutatakse järgmise valemi abil:

$$\hat{\psi}_i = \int \hat{G} d\hat{F}_i = \frac{1}{N} \left( \bar{R}_i^\psi - \frac{1}{2} \right), \quad (4)$$

kus  $\bar{R}_i^\psi$  tähistab  $i$ -nda rühma pseudo-astakute aritmeetilist keskmist ja  $\hat{F}_i$  on rühma  $i$  empiiriline jaotusfunktsioon.

**Kaalumata empiiriline segujaotus**  $\hat{G}$  saadakse  $d$  rühma empiiriliste jaotus-

funktsioonide keskmisena:

$$\hat{G} = \frac{1}{d} \sum_{i=1}^d \hat{F}_i,$$

kus iga  $\hat{F}_i$  põhineb rühma  $i$  vaatlustel  $X_{i1}, \dots, X_{in_i}$ .

**Pseudo-astakud** põhinevad fikseeritud suhtelistel efektidel ja pakuvad alternatiivi traditsioonilistele astakutele, vähendades tasakaalustamata valimi korral tekkivaid nihkeid (Brunner *et al.*, 2018; Brunner *et al.*, 2020). Pseudo-astakud defineeritakse järgmiselt:

$$R_{ik}^\psi = \frac{1}{2} + N\hat{G}(X_{ik}), \quad (5)$$

kus  $N$  on vaatluste koguarv ja

$$\hat{G}(X_{ik}) = \frac{1}{d} \sum_{r=1}^d \frac{1}{n_r} \sum_{\ell=1}^{n_r} c(X_{ik} - X_{r\ell}),$$

kus  $d$  on rühmade koguarv,  $n_r$  on rühma  $r$  suurus,  $c$  on võrdlusfunktsioon, defineeritud valemis (1).

Oluline omadus pseudo-astakutel põhineva testi puhul on see, et **võrdsete valimimahtude korral langevad astakud ja pseudo-astakud kokku**. Selle põhjuseks on see, et kui kõik rühmad on sama suurusega, siis kaalutud segujaotus  $H$  ja kaalumata segujaotus  $G$  on võrdsed:

$$H = G = \frac{1}{d} \sum_{i=1}^d F_i.$$

### **Pseudo-astakutel põhineva Kruskal–Wallise teststatistiku avaldis**

Nii nagu klassikalise Kruskal–Wallise teststatistiku saab väljendada kaalutud suhteliste efektide hinnangute  $\hat{p}_i$  kaudu, on võimalik koostada analoogne teststatistik ka pseudo-astakute põhjal. Selleks kasutatakse kaalumata suhteliste efektide hinnanguid  $\hat{\psi}_i$ .

Valemi (4) samaväärne esitus on  $\bar{R}_i^\psi = N\hat{\psi}_i + \frac{1}{2}$ . Klassikalise Kruskal–Wallise test-

statistiku valem on (3). Kui kasutada selle struktuuriga sarnast avaldist pseudo-astakute puhul, saadakse:

$$K^\psi = \frac{12}{N(N+1)} \sum_{i=1}^d n_i \left( \bar{R}_i^\psi - \frac{N+1}{2} \right)^2. \quad (6)$$

Asendades valemi (4) samaväärne esitus valemisse (6), saadakse

$$K^\psi = \frac{12N}{N+1} \sum_{i=1}^d n_i \left( \hat{\psi}_i - \frac{1}{2} \right)^2.$$

Saadud teststatistik kontrollib, kui palju erinevad rühmade kaalumata suhtelised efektid  $\hat{\psi}_i$  väärtusest  $\frac{1}{2}$ , võttes arvesse ka rühmade suurusi  $n_i$ .

### Näide: kaalumata suhteliste efektide arvutamine

Järgnevalt on toodud näide, mis illustreerib kaalumata suhteliste efektide arvutamist. Aluseks on võetud samad kolm rühma, nagu varasemas näites [Näide: kaalutud suhteliste efektide arvutamine](#).

Kui iga rühma vaatlused arvestatakse võrdselt, olenemata nende valimimahust, siis kasutatakse segujaotust  $G$ , kus iga rühma kaal on  $\frac{1}{d} = \frac{1}{3}$ . Seega  $\omega_1 = \omega_2 = \omega_3 = \frac{1}{3}$ .

Kaalumata segujaotus  $G$  kohal  $x$  on siis

$$G(x) = \frac{1}{3}F_1(x) + \frac{1}{3}F_2(x) + \frac{1}{3}F_3(x),$$

kus  $F_i$  on rühma  $i$  jaotusfunktsioon.

On teada, et  $\psi_i = P(Z < X_i) + \frac{1}{2}P(Z = X_i)$ , kus  $Z \sim G$  on juhuslik suurus,  $X_i \sim F_i$  pärineb rühma  $i$  jaotusest. Kuna tegemist on pidevate jaotustega, kehtib  $P(Z = X_i) = 0$ , mistõttu piisab leida  $\psi_i = P(Z < X_i)$ .

Kuna  $Z$  on valitud segujaotusest  $G$ , kus iga  $X_r$  on võrdset tõenäoline, kehtib:

$$\psi_i = \sum_{r=1}^3 \frac{1}{3} P(X_r < X_i),$$

kus  $X_r \sim F_r$  on vastava rühma jaotusest pärit juhuslik suurus.

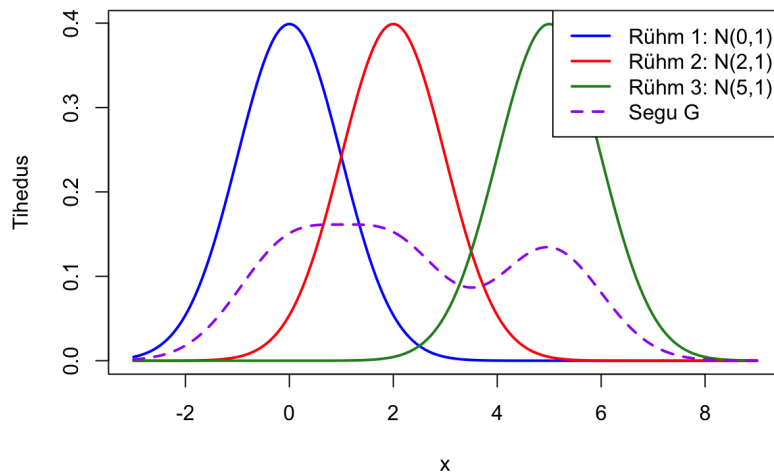
Kuna tegemist on normaaljaotustega  $\mathcal{N}(\mu_i, \sigma^2)$ , siis saab tõenäosused  $P(X_r < X_i)$  arvutada kujul:

$$P(X_r < X_i) = \Phi\left(\frac{\mu_i - \mu_r}{\sqrt{2\sigma^2}}\right).$$

Saadi järgmised kaalumata suhtelised efektid (vt lisa [Kaalumata suhteliste efektide arvutamine](#)):

$$\psi_1 \approx 0,193; \quad \psi_2 \approx 0,479; \quad \psi_3 \approx 0,828.$$

Kaalumata suhtelised efektid kirjeldavad, kui suure tõenäosusega on iga rühma vaatlus väiksem kui juhuslikult valitud vaatlus segujaotusest  $G$ . Joonisel (2) on kujutatud kolme rühma jaotused ja nende põhjal arvatud segujaotuse  $G$  tihedusfunktsioon.



Joonis 2: Kolme rühma ja segujaotuse  $G$  tihedusfunktsioon

## 2 Astakute ja pseudo-astakute arvutamise näide

Selles näites leitakse pseudo-astakud kasutades eelnevalt toodud definitsiooni ja valemi (5).

Olgu meil 3 rühma ja kokku 8 vaatlust.

$$\text{Rühm 1 : } n_1 = 3, X_{11} = 1, X_{12} = 2, X_{13} = 6;$$

$$\text{Rühm 2 : } n_2 = 2, X_{21} = 4, X_{22} = 7;$$

$$\text{Rühm 3 : } n_3 = 3, X_{31} = 3, X_{32} = 5, X_{33} = 8.$$

Kogu vaatluste arv on  $N = 8$ .

Valitud väärtused on sellised, et tavalised astakud (järjekorranumbrid) langevad kokku vaatluste väärtustega. See võimaldab keskenduda ainult pseudo-astakute arvutamisele.

### 2.1 Funktsiooni $c(u)$ väärtused

Pseudo-astakute leidmiseks arvutatakse iga vaatluse jaoks funktsiooni (1) väärtused kõikide teiste vaatluste suhtes. Allpool tuuakse üks näide (vaatlus  $X_{11} = 1$ ) ning ülejäänud arvutused on analoogsed.

1. rühm  $X_{11} = 1$

$$[c(1-X_{11}), c(1-X_{12}), c(1-X_{13}), c(1-X_{21}), c(1-X_{22}), c(1-X_{31}), c(1-X_{32}), c(1-X_{33})]$$

Nüüd valemis asendatakse juhuslikud suurused  $X_{ik}$  vaatluste väärtustega, kusjuures

$$X_{11} = 1, X_{12} = 2, X_{13} = 6, X_{21} = 4, X_{22} = 7, X_{31} = 3, X_{32} = 5, X_{33} = 8.$$

$$\begin{aligned}
& [c(1-1), c(1-2), c(1-6), c(1-4), c(1-7), c(1-3), c(1-5), c(1-8)] = \\
& [c(0), c(-1), c(-5), c(-3), c(-6), c(-2), c(-4), c(-7)] = \\
& [0,5, 0, 0, 0, 0, 0, 0]
\end{aligned}$$

Korrates sama arvutust kõigi ülejäänud vaatlustega, saadakse  $8 \times 8$  maatriks, kus iga rida vastab konkreetsele vaatlusele  $X_{ik}$  ning iga elemendi väärtus kirjeldab, kas vaatlus  $X_{ik}$  on väiksem, võrdne või suurem võrreldes vaatlustega  $X_{r\ell}$ .

$$\begin{pmatrix}
0,5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0,5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0,5 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0,5 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0,5 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0,5 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0,5 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0,5
\end{pmatrix}$$

## 2.2 Pseudo-astakute arvutamine

Allpool tuuakse näitena esimese vaatluse  $X_{11} = 1$  pseudo-astaku arvutus. Ülejäänud vaatluste puhul toimub arvutus analoogiliselt.

$R_{11}^{\psi}$  ehk vaatluse  $X_{11} = 1$  pseudo-astak

Summeeritakse  $c(1 - X_{r\ell})$  üle kõikide rühmade ( $r = 1, \dots, 3$ ) ja vaatlusindeksite ( $\ell$ ):

$$\text{Rühma 1 keskmine} = \frac{1}{n_1} [c(1-1) + c(1-2) + c(1-6)] = \frac{1}{3} [0,5 + 0 + 0] = 0,1667;$$

$$\text{Rühma 2 keskmine} = \frac{1}{n_2} [c(1-4) + c(1-7)] = \frac{1}{2} [0 + 0] = 0;$$

$$\text{Rühma 3 keskmine} = \frac{1}{n_3} [c(1-3) + c(1-5) + c(1-8)] = \frac{1}{3} [0 + 0 + 0] = 0.$$

Seega

$$R_{11}^{\psi} = \frac{1}{2} + \frac{8}{3} \times (0,1667 + 0 + 0) \approx 0,9444.$$

Analoogiliselt saab leida  $R_{12}^{\psi}, R_{13}^{\psi}, R_{21}^{\psi}, R_{22}^{\psi}, R_{31}^{\psi}, R_{32}^{\psi}, R_{33}^{\psi}$ .

Lõplikud pseudo-astakute väärtused on toodud tabelis (1).

Tabel 1: Arvutatud astakud ja pseudo-astakud

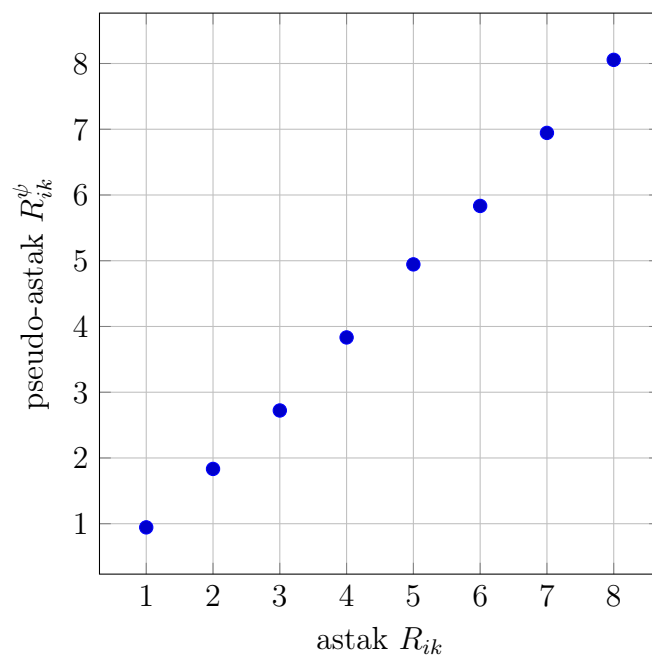
$(i, k)$	$X_{ik}$	$R_{ik}^{\psi}$
(1, 1)	1	0,9444
(1, 2)	2	1,8333
(1, 3)	6	5,8334
(2, 1)	4	3,8333
(2, 2)	7	6,9445
(3, 1)	3	2,7224
(3, 2)	5	4,9445
(3, 3)	8	8,0554

Veerus  $X_{ik}$  on esitatud vaatluste väärtused, mis on selles näites võrdsed tavaliste astakutega. Veerus  $R_{ik}^{\psi}$  on vastavad pseudo-astakud, mis võtavad kõiki rühmi arvesse võrdsete kaaludega  $\frac{1}{d}$ .

Võrdluses pseudo-astakutega ilmneb huvitav muster (vt joonis (3)): **enamiku vaatluste puhul on pseudo-astaku väärtused madalamad kui vastavad tavalised astakud**, välja arvatud suurima väärtuse  $X = 8$  puhul, mille pseudo-astak on pisut kõrgem.

Üldjuhul võivad pseudo-astakud olla sarnased klassikalistele astakutele, kuid nende väärtused võivad erineda märgatavalt, eriti järgmistel juhtudel:

- kui rühmade valimid on tasakaalustamata ehk valimimahud rühmades on erinevad;
- kui rühmade jaotusfunktsioonid on erinevad (Brunner *et al.*, 2020).



Joonis 3: Astakud vs pseudo-astakud

Pseudo-astakud saab leida kasutades **tarkvara R** (R Core Team, 2024). Selles näites on kasutatud lisapaketti **rankFD**. Lisapaketi juhend on koostatud (Konietschke *et al.*, 2019). Koodiga saab tutvuda vaadates [Lisa 2. Pseudo-astakute arvutamine tarkvara R abil](#).

### 3 Astakutel põhinevate testide ootamatud tulemused

Artiklis (Brunner *et al.*, 2020) käsitletakse olukordi, kus traditsiooniliste astakutel põhinevad testid, näiteks Kruskal–Wallise test, võivad anda eksitavaid tulemusi, kui rühmade valimid on ebavõrdsed. Sellistes olukordades võib teststatistiku väärtus sõltuda valimimahust rohkem kui tegelikust jaotuste erinevusest. See tähendab, et test võib nullhüpoteesi tagasi lükata või mitte lükata tagasi, sõltuvalt ainult rühmade suurusest, kuigi jaotuste struktuur (näiteks rühmade vahelised võrdlustõenäosused (vt valem (7)) jääb samaks.

#### 3.1 Mittetransitiivne struktuur segujaotuste korral

Käesoleva näite aluseks on artiklis (Brunner *et al.*, 2020, lk 354) toodud trikkvärin-gute näide. Artiklis vaadeldakse kolme diskreetset juhuslikku suurust  $X_i$ ,  $i = 1, 2, 3$ , mille vahel kehtib mittetransitiivne võrdlussuhe:  $P(X_1 < X_3) = P(X_3 < X_2) = P(X_2 < X_1) = \frac{7}{12}$ .

Käesolevas töös on sama idee kohandatud pidevate jaotuste jaoks, kasutades normaaljaotuste segusid. Segu komponentide keskväärtused on valitud nii, et tekiks analoogne mittetransitiivne struktuur.

Olgu antud kolm sõltumatut juhuslikku suurust  $X_i$ ,  $i = 1, 2, 3$ . Iga juhuslik suurus  $X_i$  on võrdse kaaluga segu normaaljaotustest  $\mathcal{N}(\mu_{ij}, 1)$ , mille komponentide keskväärtused on järgmised:

$$\mu_1 = \{0, 7, 8, 11, 12, 13\},$$

$$\mu_2 = \{4, 5, 6, 9, 10, 17\},$$

$$\mu_3 = \{1, 87, 1, 87, 1, 87, 15, 13, 15, 13, 15, 13\}.$$

Iga juhuslike suuruste paari  $(X_i, X_j)$  korral leitakse tõenäosus  $P(X_i < X_j)$ , kus  $X_i$

ja  $X_j$  on sõltumatud vaatlused oma vastavatest segujaotustest. Kuna tegemist ei ole tavaliste normaaljaotustega, vaid nende segudega, tuleb arvestada kõigi segukomponentide kombinatsioonidega. Eeldades, et igas  $X_i$ -s on kuus komponenti ja nende kaalud võrdsed, saab kogu tõenäosuse avaldada järgmise valemiga:

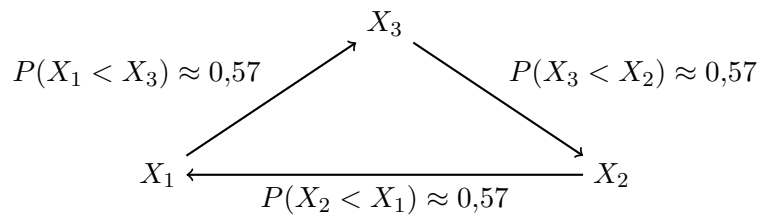
$$P(X_i < X_j) = \frac{1}{36} \sum_{k=1}^6 \sum_{\ell=1}^6 \Phi \left( \frac{\mu_{X_{j\ell}} - \mu_{X_{ik}}}{\sqrt{2}} \right), \quad (7)$$

kus  $\mu_{X_{ik}}$  ja  $\mu_{X_{j\ell}}$  tähistavad vastavalt  $X_i$  ja  $X_j$  segukomponentide keskvaartusi ning  $\Phi$  on standardse normaaljaotuse jaotusfunktsioon.

Valem (7) kirjeldab, kui suure tõenäosusega jääb juhuslikult valitud vaatlus juhuslikust suuruselt  $X_i$  väiksemaks kui vaatlus juhuslikust suuruselt  $X_j$ , võttes arvesse kõiki võimalikke segukomponentide kombinatsioone.

Põhjaliku arvutuskäiguga saab tutvuda vaadates lisa [Tõenäosuste arvutamine](#).

Pärast arvutamist saadi järgmised tulemused:  $P(X_1 < X_3) = P(X_3 < X_2) = P(X_2 < X_1) \approx 0,57$ , mis tähendab, et tekib suletud tsükkel. See näide illustreerib **mittetransitiivsust**, mida artikli autorid (Brunner *et al.*, 2020) seostavad astakupõhiste testide üllatavate käitumismustritega.



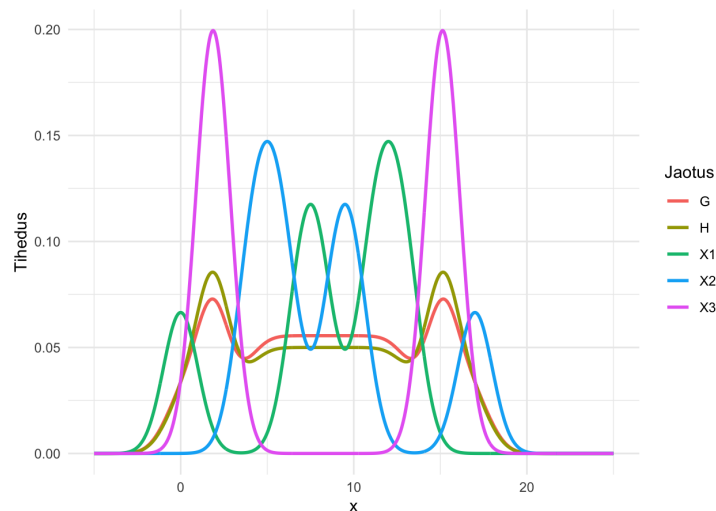
### 3.2 Simulatsioonid: klassikalise ja pseudo-astakutel põhineva Kruskal–Wallise testi võrdlus

Edasi uuritakse simulatsioonide abil, kuidas käituvad klassikaline Kruskal–Wallise test ja pseudo-astakutel põhinev test olukorras, kus andmete aluseks on mittetransitiivne struktuur. Kuigi rühmade jaotused ei ole identsed, on võrdlustõenäosused

igas rühmapaaris väga sarnased ( $P(X_i < X_j) \approx 0,57$ ), mistõttu võiks eeldada, et testid ei tuvasta olulist erinevust ehk jääb kehtima  $H_0$  : rühmade jaotused on samad.

Esmalt tehti võrdlus ühe juhuslikult genereeritud valimi korral, et illustreerida testi tulemusi konkreetses andmestikus. Seejärel kordasid simulatsioonid samu arvutusi 10000 korral, et võrrelda testide käitumist erinevate valimimahu kombinatsioonide puhul.

Kõik  $X_i$  on võrdselt kaalutud segud  $\mathcal{N}(\mu_{ij}, 1)$ , keskväärtustega nagu eespool. Joonisel (4) on esitatud iga rühma tihedusfunktsioon ning lisatud kaalutud segujaotus  $H$  ja kaalumata segujaotus  $G$ .



Joonis 4: Rühmade  $X_1$ ,  $X_2$ ,  $X_3$  ja nende segujaotuste  $H$  (kaalutud) ja  $G$  (kaalumata) tihedusfunktsioonid

Simulatsioonides kasutati kolme erinevat valimimahu kombinatsiooni:

- võrdsed: (10, 10, 10);
- ebavõrdsed 1: (3, 18, 9);
- ebavõrdsed 2: (8, 16, 6).

Tabelis (2) on esitatud kaalutud ( $p_i$ ) ja kaalumata ( $\psi_i$ ) suhteliste efektide väärtused iga simulatsioonis kasutatud valimimahu kombinatsiooni korral (vt koodi [Kaalumata ja kaalutud suhteliste efektide arvutamine](#)).

Tabel 2: Kaalutud ja kaalumata suhtelised efektid

Valimid	$p_1$	$p_2$	$p_3$	$\psi_1$	$\psi_2$	$\psi_3$
10, 10, 10	0,500	0,500	0,500	0,500	0,500	0,500
3, 18, 9	0,520	0,510	0,470	0,500	0,500	0,500
8, 16, 6	0,520	0,490	0,480	0,500	0,500	0,500

Kuigi jaotused on erinevad, kehtib kõigis paarides  $P(X_1 < X_3) = P(X_3 < X_2) = P(X_2 < X_1) \approx 0,57$  (vt lisa [Tõenäosuste arvutamine simulatsioonide abil](#)).

Sellises olukorras võiks oodata, et test ei tuvasta erinevust. Järgmised tulemused näitavad, et klassikalise Kruskal–Wallise testi järeldus võib sõltuda rühmade valimimahust.

### Üksiku simulatsiooni tulemused

Ühekordse simulatsiooni korral saadi järgmised tulemused:

- **võrdsed valimid:**  $p$ -väärtus oli suur ( $p > 0,1$ ), mistõttu jäädi nullhüpoteesi juurde, st rühmade jaotused on sarnased;
- **ebavõrdsed valimid 1 ja 2:**  $p$ -väärtus tuli alla 0,1, nullhüpotees lükati tagasi, st vähemalt üks jaotus erineb.

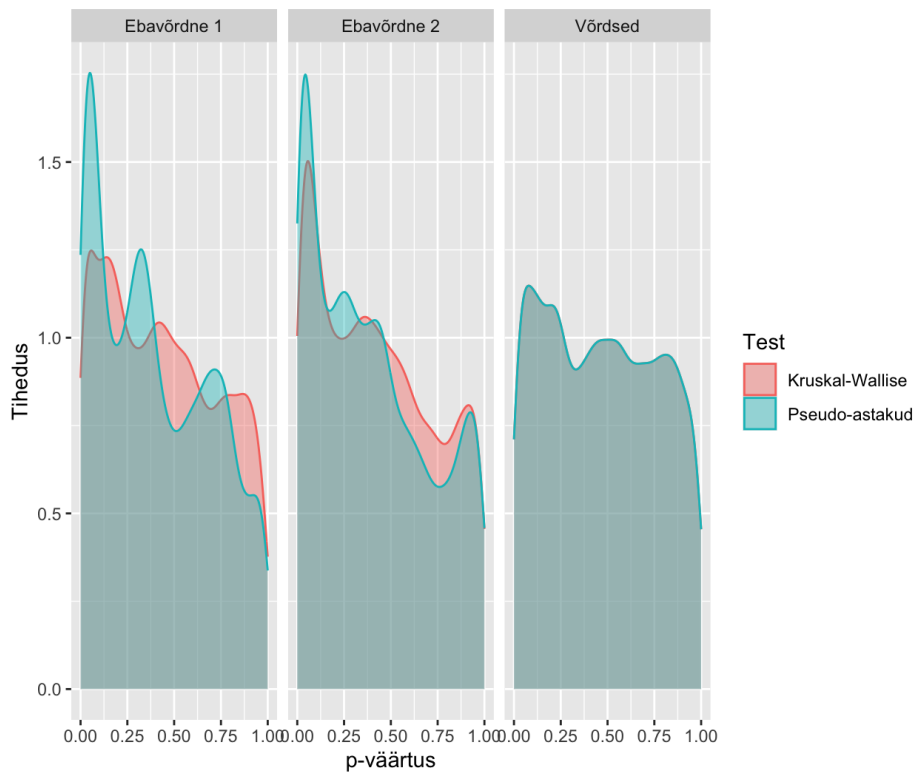
### Simulatsiooni tulemused väikeste valimitega

Esimese simulatsiooni korral viidi iga test läbi 10000 korral valimimahu kombinatsioonidega nagu enne. Iga kord arvutati  $p$ -väärtus nii klassikalise kui ka pseudoastakutel põhineva Kruskal–Wallise testile. Tabelis (3) on esitatud protsent kui sa-

geli nullhüpoteesi lükati tagasi (ehk väideti, et rühmade jaotused erinevad) olulisuse nivooga  $\alpha = 0,1$ .

Tabel 3: Nullhüpoteesi tagasilükkamise sagedus väiksete valimite korral (%)

Valimid	Kruskal–Wallise test	Pseudo-astakutel põhinev test
10, 10, 10	12,89	12,89
3, 18, 9	15,21	22,44
8, 16, 6	18,59	22,62



Joonis 5:  $p$ -väärtuste jaotused väikeste valimite korral

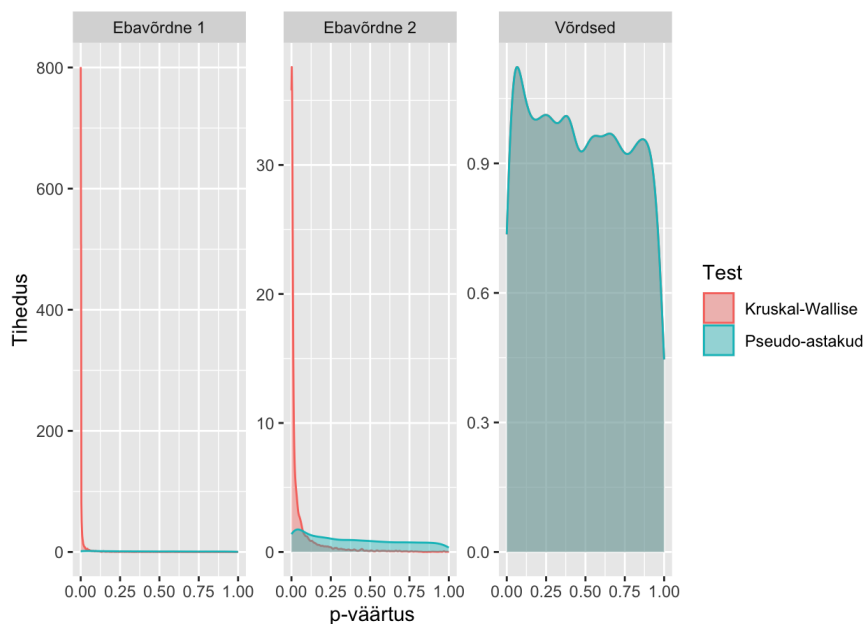
Mõlema testi puhul on  $H_0$  tagasilükkamise sagedus võrdsete valimite korral sama. Ebavõrdsete valimite korral lükkab Kruskal–Wallise test nullhüpoteesi tagasi harvemini kui pseudo-astakutel põhinev test. Joonis (5) illustreerib  $p$ -väärtuste tihedusjaotusi Kruskal–Wallise ja pseudo-astakutel põhineva testi korral.

## Simulatsiooni tulemused suurte valimitega

Järgmisena korrati sama simulatsiooni suurendatud valimitega (iga rühma suurus korrutatud 100-ga):

Tabel 4: Nullhüpoteesi tagasilükkamise sagedus suurte valimite korral (%)

Valimid	Kruskal–Wallise test	Pseudo-astakutel põhinev test
1000, 1000, 1000	13,23	13,23
300, 1800, 900	96,96	21,81
800, 1600, 600	83,86	22,37



Joonis 6:  $p$ -väärtuste jaotused suurte valimite korral

Tulemused tabelis (4) näitavad, et kui valimid on suured ja tasakaalustamata, kaldub klassikaline Kruskal–Wallise test  $H_0$  tagasi lükkama väga suure sagedusega. See juhtub, kuigi rühmade võrdlustõenäosused on sarnased. Kruskal–Wallise testil koguneb rohkem  $p$ -väärtusi alla kriitilise piiri, samas kui pseudo-astakutel põhinev test annab ühtlasema jaotuse (vt joonis (6)).

Tulemused näitavad, et klassikaline Kruskal–Wallise test võib anda eksitava tulemuse, kuna teststatistikut mõjutavad tugevalt valimimahud, mitte ainult jaotuste kuju. See probleem ilmneb eriti siis, kui valimid on tasakaalustamata.

Pseudo-astakutel põhinev test seevastu säilitab stabiilsema käitumise, sest tema suhtelised efektid ei sõltu valimimahust.

Arvestades, et pseudo-astakutel põhinev test tugineb kaalumata suhtelistele efektidele  $\psi_i$ , saab selle testi eesmärki väljendada järgmiselt: test kontrollib, kas iga rühma suhteline efekt erineb väärtusest  $\frac{1}{2}$ . Nullhüpoteesi saab sel juhul sõnastada kujul

$$H_0 : \psi_1 = \psi_2 = \psi_3 = \frac{1}{2}.$$

Kuna  $\psi_i$  väärtused sõltuvad ainult rühmade jaotustest  $F_i$  ega ole mõjutatud valimimahust, on see hüpotees defineeritud juba enne andmete kogumist. Seevastu kaalutud suhteliste efektide  $p_i$  väärtused sõltuvad ka valimi struktuurist, mistõttu selguvad need alles pärast valimi võtmist. Sellest tulenevalt ei saa koostada analoogset nullhüpoteesi kujul  $H_0 : p_1 = p_2 = p_3 = \frac{1}{2}$ .

## Kokkuvõte

Käesolevas töös uuriti pseudo-astakute kasutamist mitme rühma võrdlemisel ning võrreldi sellel põhineva testi käitumist klassikalise Kruskal–Wallise testiga. Töö eesmärk oli selgitada pseudo-astakute omadusi ning näidata, millistes olukordades need annavad usaldusväärsemaid tulemusi kui tavalistel astakutel põhinevad testid.

Metoodikaks kasutati simulatsioone, kus andmeid genereeriti normaaljaotuste segudena. Katsetati erinevaid valimimahu kombinatsioone, nii võrdselt jaotatud kui ka tasakaalustamata. Iga simulatsiooni puhul võrreldi Kruskal–Wallise testi ja pseudo-astakutel põhineva testi p-väärtuste jaotust ning nullhüpoteesi tagasilükkamisesagedust.

Tulemused näitasid, et kui rühmade jaotused on keerulised (näiteks normaaljaotuste segud) ja valimimahud ebavõrdsed, võib klassikaline Kruskal–Wallise test anda eksitavaid järeldusi. Nimelt võib test nullhüpoteesi sagedamini tagasi lükata mitte seetõttu, et jaotused oleksid sisuliselt erinevad, vaid kuna teststatistik on tundlik valimimahtude ebavõrdsusele. See probleem avaldus eriti hästi mittetransitiivsuuse näites, kus kõik rühmadevahelised võrdlustõenäosused olid ligikaudu võrdsed ( $P(X_1 < X_3) = P(X_3 < X_2) = P(X_2 < X_1) \approx 0,57$ ), kuid Kruskal–Wallise test andis erineva järelduse sõltuvalt valimimahust. Pseudo-astakutel põhinev test käitus sellistes olukordades stabiilsemalt. Kuna see test kasutab suhtelisi efekte, mis ei sõltu rühmade valimimahust, püsis selle teststatistik ja p-väärtuste jaotus tasakaalustatult ka siis, kui valimimahud olid väga erinevad.

Kokkuvõttes kinnitasid tulemused, et pseudo-astakutel põhinev lähenemine võimaldab mitme rühma võrdluses teha usaldusväärsemaid järeldusi olukordades, kus klassikalise astakupõhise testi eeldused ei ole täidetud. Selliseid olukordi esineb praktikas sageli, eriti kui andmed ei ole pärit ühest lihtsast jaotusest või kui valimimahud rühmades on ebavõrdsed. Seetõttu võib pseudo-astakutel põhinevat testi pidada oluliseks tööriistaks mitteparameetriliste meetodite arsenalis.

## Kasutatud allikad

- Brunner, E., F. Konietschke, Arne C. Bathke ja M. Pauly (2018). “Ranks and Pseudo-ranks — Paradoxical Results of Rank Tests”, lk. 1–19. DOI: [10.48550/arXiv.1802.05650](https://doi.org/10.48550/arXiv.1802.05650).
- (2020). “Ranks and Pseudo-ranks — Surprising Results of Certain Rank Tests in Unbalanced Designs”. *International Statistical Review* 89.2, lk. 349–366. DOI: [10.1111/insr.12418](https://doi.org/10.1111/insr.12418).
- Konietschke, Frank, Sarah Friedrich, Edgar Brunner ja Markus Pauly (2019). *rankFD: Rank-Based Tests for General Factorial Designs*. <https://CRAN.R-project.org/package=rankFD>. R package manual.
- OpenAI ChatGPT (2025). *Vestluspõhine abiline statistiliste simulatsioonide ja jooniste juures*. Vestlus OpenAI ChatGPT-ga, kasutaja Mihhail Bobeško, veebruar - mai 2025. URL: <https://chat.openai.com>.
- R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

## Lisa 1. Suhteliste efektide arvutamine

### Kaalutud suhteliste efektide arvutamine

On teada, et  $P(X_i < X_i) = \frac{1}{2}$ , seega

$$P(X_1 < X_1) = \frac{1}{2}, \quad P(X_2 < X_2) = \frac{1}{2}, \quad P(X_3 < X_3) = \frac{1}{2}.$$

Järgmiseks leitakse erinevate rühmade tõenäosused  $P(X_r < X_i)$ , kui  $r \neq i$  ehk  $P(X_1 < X_2), P(X_2 < X_1), P(X_1 < X_3), P(X_3 < X_1), P(X_2 < X_3)$  ja  $P(X_3 < X_2)$ .

$$P(X_1 < X_2) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{2}}\right) = \Phi\left(\frac{2 - 0}{\sqrt{2}}\right) = \Phi(1,41) \approx 0,92073;$$

$P(X_2 < X_1)$  on  $P(X_1 < X_2)$  vastandtõenäosus:

$$P(X_2 < X_1) = 1 - P(X_1 < X_2) = 1 - 0,92073 = 0,07927;$$

Analoogiliselt:

$$P(X_1 < X_3) = \Phi\left(\frac{\mu_3 - \mu_1}{\sqrt{2}}\right) = \Phi\left(\frac{5 - 0}{\sqrt{2}}\right) = \Phi(3,54) \approx 0,99980$$

$$P(X_3 < X_1) = 1 - P(X_1 < X_3) = 1 - 0,99980 = 0,0002$$

$$P(X_2 < X_3) = \Phi\left(\frac{\mu_3 - \mu_2}{\sqrt{2}}\right) = \Phi\left(\frac{5 - 2}{\sqrt{2}}\right) = \Phi(2,12) \approx 0,98300$$

$$P(X_3 < X_2) = 1 - P(X_2 < X_3) = 1 - 0,98300 = 0,017$$

Nüüd nende väärtuste abil saab leida suhtelised efektid  $p_1, p_2$  ja  $p_3$  :

$$p_1 = \omega_1 P(X_1 < X_1) + \omega_2 P(X_2 < X_1) + \omega_3 P(X_3 < X_1)$$

$$p_1 = 0,3 \cdot 0,5 + 0,3 \cdot 0,07927 + 0,4 \cdot 0,0002 \approx 0,174;$$

$$p_2 = \omega_1 P(X_1 < X_2) + \omega_2 P(X_2 < X_2) + \omega_3 P(X_3 < X_2)$$

$$p_2 = 0.3 \cdot 0.92073 + 0.3 \cdot 0.5 + 0.4 \cdot 0.017 \approx 0.433;$$

$$p_3 = \omega_1 P(X_1 < X_3) + \omega_2 P(X_2 < X_3) + \omega_3 P(X_3 < X_3)$$

$$p_3 = 0.3 \cdot 0.9998 + 0.3 \cdot 0.983 + 0.4 \cdot 0.5 \approx 0.795.$$

Joonis (1) on tehtud tarkava R abil:

```
#Rühm 1: mu=0, sigma=1
#Rühm 2: mu=2, sigma=1
#Rühm 3: mu=5, sigma=1
#Kaalud
w1 <- 0.3
w2 <- 0.3
w3 <- 0.4
x <- seq(-3, 9, length.out = 300)
#Tihedus
y1 <- fun(x, 0, 1)
y2 <- fun(x, 2, 1)
y3 <- fun(x, 5, 1)

#Segujaotuse tihedus:
yH <- w1*y1 + w2*y2 + w3*y3

plot(x, y1, type = "l", col = "blue", lwd = 2,
      xlab = "x", ylab = "Tihedus",
      ylim = c(0, max(y1, y2, y3, yH)))
```

```

lines(x, y2, col = "red", lwd = 2)
lines(x, y3, col = "forestgreen", lwd = 2)
lines(x, yH, col = "black", lwd = 2, lty = 2)

legend("topright",
      legend = c("Rühm 1: N(0,1)",
                 "Rühm 2: N(2,1)",
                 "Rühm 3: N(5,1)",
                 "Segu H"),
      col = c("blue", "red", "forestgreen", "black"),
      lwd = 2, lty = c(1,1,1,2))

```

## Kaalumata suhteliste efektide arvutamine

Teada on, et  $P(X_i < X_i) = \frac{1}{2}$ , seega:

$$P(X_1 < X_1) = P(X_2 < X_2) = P(X_3 < X_3) = \frac{1}{2}.$$

Järgmiselt leitakse erinevate rühmade tõenäosused:

$$P(X_1 < X_2) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{2}}\right) = \Phi(1,41) \approx 0,92073;$$

$$P(X_2 < X_1) = 1 - 0,92073 = 0,07927;$$

$$P(X_1 < X_3) = \Phi\left(\frac{\mu_3 - \mu_1}{\sqrt{2}}\right) = \Phi(3,54) \approx 0,99980;$$

$$P(X_3 < X_1) = 1 - 0,99980 = 0,0002;$$

$$P(X_2 < X_3) = \Phi\left(\frac{\mu_3 - \mu_2}{\sqrt{2}}\right) = \Phi(2,12) \approx 0,98300;$$

$$P(X_3 < X_2) = 1 - 0,98300 = 0,017.$$

Nüüd saab leida kaalumata suhtelised efektid  $\psi_1$ ,  $\psi_2$  ja  $\psi_3$ :

$$\psi_1 = \omega_1 P(X_1 < X_1) + \omega_2 P(X_2 < X_1) + \omega_3 P(X_3 < X_1)$$

$$\psi_1 = \frac{1}{3} \cdot 0,5 + \frac{1}{3} \cdot 0,07927 + \frac{1}{3} \cdot 0,0002 \approx 0,193;$$

Analoogiliselt

$$\psi_2 = \frac{1}{3} \cdot 0,92073 + \frac{1}{3} \cdot 0,5 + \frac{1}{3} \cdot 0,017 \approx 0,479;$$

$$\psi_3 = \frac{1}{3} \cdot 0,99980 + \frac{1}{3} \cdot 0,98300 + \frac{1}{3} \cdot 0,5 \approx 0,828.$$

Joonis (2) on tehtud analoogiliselt joonisele (1).

#Segujaotuse G tihedus

yG=(y1+y2+y3)/3

## Lisa 2. Pseudo-astakute arvutamine tarkvara R abil

```
library(rankFD)

# i=1 (3 vaadet), i=2 (2 vaadet), i=3 (3 vaadet)
X <- c(1,2,6, 4,7, 3,5,8)

ryhm <- factor(c(rep(1,3), rep(2,2), rep(3,3)))

psr(X ~ ryhm, data=NULL)
```

Tulemus:

	X	ryhm	pseudorank
1	1	1	0.9444444
2	2	1	1.8333333
3	6	1	5.8333333
4	4	2	3.8333333
5	7	2	6.9444444
6	3	3	2.7222222
7	5	3	4.9444444
8	8	3	8.0555556

## Lisa 3. Testide ootamatud tulemused: pidevate jaotustega näide

### Tõenäosuste arvutamine

Alljärgnevalt leitakse tõenäosus  $P(X_1 < X_3)$ . Ülejäänud tõenäosused  $P(X_2 < X_1)$  ja  $P(X_3 < X_2)$  saab analoogiliselt arvutada.

Tabelis on arvutatud  $\Delta_{ij} = \mu_{X_3} - \mu_{X_1}$  vahed:

$X_3 \setminus X_1$	0	7	8	11	12	13
1,87	1,87	-5,13	-6,13	-9,13	-10,13	-11,13
1,87	1,87	-5,13	-6,13	-9,13	-10,13	-11,13
1,87	1,87	-5,13	-6,13	-9,13	-10,13	-11,13
15,13	15,13	8,13	7,13	4,13	3,13	2,13
15,13	15,13	8,13	7,13	4,13	3,13	2,13
15,13	15,13	8,13	7,13	4,13	3,13	2,13

Kuna  $X_3 - X_1 \sim \mathcal{N}(\Delta_{ij}, \sigma_{\Delta}^2)$ , saab tõenäosuse avaldada kui  $P(X_1 < X_3) = \Phi\left(\frac{\Delta_{ij}}{\sqrt{2}}\right)$ . Tõenäosuse leidmiseks kasutatakse tarkvara R.

```
X1 <- c(0, 7, 8, 11, 12, 13)
X3 <- c(1.87, 1.87, 1.87, 15.13, 15.13, 15.13)
sigma_delta <- sqrt(2)
Delta <- outer(X3, X1, "-")
#P(X1 < X3) = Phi(Delta / sigma_delta)
prob <- pnorm(Delta / sigma_delta)
x1_x3 <- mean(prob)
x1_x3
```

$$P(X_1 < X_3) = \frac{1}{36} \sum_{i=1}^6 \sum_{j=1}^6 \Phi\left(\frac{\mu_{X_3} - \mu_{X_1}}{\sqrt{2}}\right) \approx 0,57.$$

## Simulatsioonid

```
set.seed(1234)
#Keskväärtused
F1 <- c(0,7,8,11,12,13)
F2 <- c(4,5,6,9,10,17)
F3 <- c(1.87,1.87,1.87,15.13,15.13,15.13)
#Standardhälve
std <- 1
```

Funktsioon andmete genereerimiseks normaaljaotuste segust:

```
normaaljaotuse_segu <- function(keskv, suurus) {
  komponendid <- sample(1:length(keskv), size = suurus, replace = TRUE)
  prov <- sapply(komponendid, function(i) rnorm(1, mean = keskv[i], sd = std))
  return(prov)
}
```

Võrdsete valimimahtudega andmete genereerimine:

```
n_vordne <- c(10, 10, 10)
vordne <- data.frame(
  Value = c(normaaljaotuse_segu(F1, n_vordne[1]),
            normaaljaotuse_segu(F2, n_vordne[2]),
            normaaljaotuse_segu(F3, n_vordne[3])),
  Group = factor(rep(c("F1", "F2", "F3"), times = n_vordne))
)
#Test
kruskal.test(Value ~ Group, data = vordne)
```

Tulemuseks on:

```
> kruskal.test(Value ~ Group, data = vordne)
Kruskal-Wallis rank sum test
data: Value by Group
Kruskal-Wallis chi-squared = 2.8413, df = 2, p-value = 0.2416
```

Kuna  $p$ -väärtus on suurem kui 0,1, jäädi nullhüpoteesi juurde.

Ebavõrdsete valimimahtudega andmete genereerimine:  $n_1 = 3, n_2 = 18, n_3 = 9$

```

n_ebav <- c(3, 18, 9)
ebav <- data.frame(
  Value = c(normaaljaotuse_seg(F1, n_ebav[1]),
            normaaljaotuse_seg(F2, n_ebav[2]),
            normaaljaotuse_seg(F3, n_ebav[3])),
  Group = factor(rep(c("F1", "F2", "F3"), times = n_ebav))
)
#Test
kruskal.test(Value ~ Group, data = ebav)

```

Tulemuseks on:

```

> kruskal.test(Value ~ Group, data = ebav)
Kruskal-Wallis rank sum test
data: Value by Group
Kruskal-Wallis chi-squared = 16.03, df = 2, p-value = 0.0003305

```

Kuna  $p$ -väärtus on väiksem kui 0,1, võeti vastu alternatiivne hüpotees.

Ebavõrdsete valimimahtudega andmete genereerimine 2:  $n_1 = 8, n_2 = 16, n_3 = 6$

```

n_ebav2 <- c(8, 16, 6)
ebav2 <- data.frame(
  Value = c(normaaljaotuse_seg(F1, n_ebav2[1]),
            normaaljaotuse_seg(F2, n_ebav2[2]),
            normaaljaotuse_seg(F3, n_ebav2[3])),
  Group2 = factor(rep(c("F1", "F2", "F3"), times = n_ebav2))
)
#Test
kruskal.test(Value ~ Group2, data = ebav2)

```

Tulemuseks on:

```

> kruskal.test(Value ~ Group2, data = ebav2)
Kruskal-Wallis rank sum test
data: Value by Group2
Kruskal-Wallis chi-squared = 6.2489, df = 2, p-value = 0.04396

```

Kuna  $p$ -väärtus on väiksem kui 0,1, võeti vastu alternatiivne hüpotees.

## Tõenäosuste arvutamine simulatsioonide abil

On ka teine võimalus tõenäosuste arvutamiseks kasutades tehtud simulatsioonid.

Genereeritakse juhuslikud proovid:

```
simulatsiooni_suurus <- 100000
X1 <- normaaljaotuse_seg(F1, simulatsiooni_suurus)
X2 <- normaaljaotuse_seg(F2, simulatsiooni_suurus)
X3 <- normaaljaotuse_seg(F3, simulatsiooni_suurus)
```

Tõenäosuste arvutamine:

```
arvuta_tõenäosus_outer <- function(A, B, chunk_size = 5000) {
  n_A <- length(A)
  n_B <- length(B)
  loendur <- 0
  total_pairs <- as.numeric(n_A) * as.numeric(n_B)
  for (i in seq(1, n_A, by = chunk_size)) {
    A_chunk <- A[i:min(i + chunk_size - 1, n_A)] #Võetakse osa A-st
    for (j in seq(1, n_B, by = chunk_size)) {
      B_chunk <- B[j:min(j + chunk_size - 1, n_B)] #Võetakse osa B-st
      loendur <- loendur + sum(outer(A_chunk, B_chunk, "<"))
    }
  }
  #Lõplik tõenäosus
  return(loendur / total_pairs)
}

w21 <- arvuta_tõenäosus_outer(X2, X1) #P(X2 < X1)
w13 <- arvuta_tõenäosus_outer(X1, X3) #P(X1 < X3)
w32 <- arvuta_tõenäosus_outer(X3, X2) #P(X3 < X2)
```

Tulemuseks on:

```
> w21
[1] 0.5686571
> w13
[1] 0.5693692
> w32
[1] 0.5691264
```

Ümardades saadud tulemused, klapiivad need enne arvutatud tõenäosustega.

## Kaalumata ja kaalutud suhteliste efektide arvutamine

```
F_ <- list(F1, F2, F3)
pij <- function(r, i) {
  mean(sapply(i, function(i) {
    mean(pnorm((i - r) / sqrt(2)))
  })))
}
#P(X_r < X_i)
P <- matrix(0, 3, 3)
for (i in 1:3) {
  for (j in 1:3) {
    if (i == j) {
      P[i, j] <- 0.5
    } else {
      P[i, j] <- pij(F_[[j]], F_[[i]])
    }
  }
}
tulemused <- lapply(names(valimid_list), function(name) {
  w <- valimid_list[[name]]
  omega <- w / sum(w)
  p_i <- as.vector(P %*% omega)
  psi_i <- colMeans(P)
  data.frame(
    Valimid = name,
    p1 = round(p_i[1], 2),
    p2 = round(p_i[2], 2),
    p3 = round(p_i[3], 2),
    psi1 = round(psi_i[1], 2),
    psi2 = round(psi_i[2], 2),
    psi3 = round(psi_i[3], 2)
  )
})
do.call(rbind, tulemused)
```

## Lisa 4. Klassikalise Kruskal-Wallise testi ja pseudo-astakutel põhineva Kruskal-Wallise testi võrdlemine

### Testide võrdlemine väiksete valimimahtude korral

```
library(rankFD)
set.seed(1234)
# Keskväärtused
F1 <- c(0,7,8,11,12,13)
F2 <- c(4,5,6,9,10,17)
F3 <- c(1.87,1.87,1.87,15.13,15.13,15.13)
std <- 1
```

```
valimid_list <- list(
  "Võrdsed"      = c(10, 10, 10),
  "Ebavõrdne 1" = c(3, 18, 9),
  "Ebavõrdne 2" = c(8, 16, 6)
)
```

Funktsioon andmete genereerimiseks:

```
normaaljaotuse_segu <- function(keskv, suurus) {
  komponendid <- sample(1:length(keskv), size = suurus, replace = TRUE)
  sapply(komponendid, function(i) rnorm(1, mean = keskv[i], sd = std))
}
simulatsioonide_arv <- 10000
```

Funktsioon ühe simulatsiooni tegemiseks:

```
simulatsioon <- function(n_vec) {
  andmed <- data.frame(
    Value = c(normaaljaotuse_segu(F1, n_vec[1]),
              normaaljaotuse_segu(F2, n_vec[2]),
              normaaljaotuse_segu(F3, n_vec[3])),
    Group = factor(rep(c("F1", "F2", "F3"), times = n_vec))
  )
}
```

```

)
#Kruskal-Wallise test
p_kw <- kruskal.test(Value ~ Group, data = andmed)$p.value
#Pseudo-astakutel põhinev test
rankfd_res <- rankFD(Value ~ Group, data = andmed)
p_pseudo <- rankfd_res$Kruskal.Wallis.Test["Group", "p-Value"]
return(c(p_kw, p_pseudo))
}
tulemused <- list()

for (i in 1:length(valimid_list)) {
  n_vec <- valimid_list[[i]]
  pvaart <- replicate(simulatsioonide_arv, simulatsioon(n_vec))

  tulemused[[i]] <- list(
    kruskal = pvaart[1, ],
    pseudo = pvaart[2, ]
  )
}

```

Nullhüpoteesi tagasilükkamine:

```

alpha <- 0.1
for (i in 1:3) {
  cat("Valimid:", paste(valimid_list[[i]], collapse = "-"), "\n")
  cat("Kruskal-Wallise testi H0 tagasilükkamine (%):", mean(tulemused[[i]]$kruskal < alpha) * 100, "\n")
  cat("Pseudo-astaku testi H0 tagasilükkamine (%): ", mean(tulemused[[i]]$pseudo < alpha) * 100, "\n\n")
}

```

Saadud tulemused:

```

Valimid: 10-10-10
Kruskal-Wallise testi H0 tagasilükkamine (%): 12.97
Pseudo-astaku testi H0 tagasilükkamine (%): 12.97

```

```

Valimid: 3-18-9
Kruskal-Wallise testi H0 tagasilükkamine (%): 14.93
Pseudo-astaku testi H0 tagasilükkamine (%): 21.84

```

```

Valimid: 8-16-6

```

Kruskal-Wallise testi H0 tagasilükkamine (%): 17.87

Pseudo-astaku testi H0 tagasilükkamine (%): 21.97

Joonis (5):

```
library(ggplot2)
library(dplyr)
p_all1 <- bind_rows(
  data.frame(p_value = tulemused[[1]]$kruskal, Test = "Kruskal-Wallise", Valimid = "Võrdsed"),
  data.frame(p_value = tulemused[[1]]$pseudo, Test = "Pseudo-astakud", Valimid = "Võrdsed"),
  data.frame(p_value = tulemused[[2]]$kruskal, Test = "Kruskal-Wallise", Valimid = "Ebavõrdne 1"),
  data.frame(p_value = tulemused[[2]]$pseudo, Test = "Pseudo-astakud", Valimid = "Ebavõrdne 1"),
  data.frame(p_value = tulemused[[3]]$kruskal, Test = "Kruskal-Wallise", Valimid = "Ebavõrdne 2"),
  data.frame(p_value = tulemused[[3]]$pseudo, Test = "Pseudo-astakud", Valimid = "Ebavõrdne 2")
)
ggplot(p_all1, aes(x = p_value, fill = Test, color = Test)) +
  geom_density(alpha = 0.4) +
  facet_wrap(~ Valimid) +
  labs(x = "p-väärtus", y = "Tihedus")
```

Joonis (4):

```
segu_tihedus <- function(x, kesk, sd = 1) {
  mean(sapply(kesk, function(mu) dnorm(x, mean = mu, sd = sd)))
}
x_v <- seq(-5, 25, length.out = 1000)
df_tihedused <- data.frame(
  x = rep(x_v, 3),
  tih = c(
    sapply(x_v, segu_tihedus, kesk = F1),
    sapply(x_v, segu_tihedus, kesk = F2),
    sapply(x_v, segu_tihedus, kesk = F3)
  ),
  Segu = rep(c("mu1", "mu2", "mu3"), each = length(x_v))
)
ggplot(df_tihedused, aes(x = x, y = tih, color = Segu)) +
  geom_line(size = 1) +
  labs(x = "x", y = "Tihedus")
```

## Testide võrdlemine suurte valimimahtude korral

Kood on analoogne, tuleb vaid asendada

```
valimid_list <- list(  
  "Võrdsed"      = c(1000, 1000, 1000),  
  "Ebavõrdne 1" = c(300, 1800, 900),  
  "Ebavõrdne 2" = c(800, 1600, 600)  
).
```

Saadud tulemused:

Valimid: 1000-1000-1000

Kruskal-Wallise testi H0 tagasilükkamine (%): 13.23

Pseudo-astaku testi H0 tagasilükkamine (%): 13.23

Valimid: 300-1800-900

Kruskal-Wallise testi H0 tagasilükkamine (%): 96.96

Pseudo-astaku testi H0 tagasilükkamine (%): 21.81

Valimid: 800-1600-600

Kruskal-Wallise testi H0 tagasilükkamine (%): 83.86

Pseudo-astaku testi H0 tagasilükkamine (%): 22.37

Joonis (6) on analoogne joonisele (5).

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Mihhail Bobeško,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Astakute ja pseudo-astakute võrdlus Kruskal-Wallise testi korral, mille juhendaja on Anne Selart, 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.

Mihhail Bobeško

14.05.2025