

TARTU ÜLIKOOL

Matemaatika ja statistika instituut

Kristjan-Martin Kirjanen

# Mitteparameetrilised testid rohkem kui kahe grupi järjestuse kontrollimiseks

Matemaatilise statistika õppekava

Bakalaureusetöö (9 EAP)

Juhendaja: Anne Selart

Tartu 2018

**Mitteparameetrilised testid rohkem kui kahe grupi järjestuse  
kontrollimiseks**

Bakalaureusetöö

Kristjan-Martin Kirjanen

Bakalaureusetöö koosneb kolmest osast. Neist esimeses kirjeldatakse Jonckheere-Terpstra testi ja leitakse teststatistiku täpne ja asümptootiline jaotus. Jonckheere-Terpstra test on kõige levinum mitteparameetriline test rohkem kui kahe järjestatud grupi võrdlemiseks. Töö teises osas kirjeldatakse üldiselt veel nelja mitteparameetrilist testi, milleks on modifitseeritud Jonckheere-Terpstra test, Terpstra-Mageli test, Cuzicki test ja Shani, Youngi ja Kangi poolt välja pakutud test. Töö kolmandas osas võrreldakse eelnevalt nimetatud viie testi võimsusi ja statistiku arvutamise algoritmi keerukusi. Testi võimsusi võrreldakse andmete simuleerimise abil.

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

**Märksõnad:** statistilised meetodid, mitteparameetrilised meetodid, Monte Carlo meetodid, statistiline kontroll

**Nonparametric tests for comparing more than two ordered groups**

Bachelor thesis

Kristjan-Martin Kirjanen

Bachelor thesis consists of three parts. First part focuses on Jonckheere-Terpstra test and finding Jonckheere-Terpstra statistics exact and asymptotic distributions. Jonckheere-Terpstra test is most used nonparametric test for comparing more than

two ordered groups. In the second part of the thesis four other nonparametric tests for comparing more than two ordered groups are briefly described. The last part of the thesis compares the power and the complexity of algorithms for calculating test statistics of all five tests.

**CERCS reasearch specialisation:** P160 Statistics, operations research, programming, actuarial mathematics

**Keywords:** statistical methods, nonparametric methods, Monte Carlo methods, statistical test

# Sisukord

Sissejuhatus	6
<b>1 Jonckheere-Terpstra test</b>	<b>7</b>
1.1 Hüpoteesid ja eeldused	7
1.2 Jonckheere'i test	8
1.3 Terpstra statistik	10
1.4 Jonckheere'i ja Terpstra statistikute täpne jaotus nullhüpoteesi kehtides, ilma võrdsete valimi väärtusteta	11
1.5 Jonckheere'i ja Terpstra statistiku keskvärtus ja dispersioon nullhüpoteesi kehtides	13
1.6 Asümptootiline jaotus	14
1.7 Võrdsed väärtused	16
<b>2 Veel teste rohkem kui kahe grupi järjestuse kontrollimiseks</b>	<b>19</b>
2.1 Modifitseeritud Jonckheere-Terpstra test	19
2.2 Terpstra-Mageli test	20
2.3 Cuzicki test	21

2.4	Shan-Young-Kangi test . . . . .	22
<b>3</b>	<b>Kirjeldatud testide võrdlemine</b>	<b>23</b>
3.1	Kirjeldatud testide ajaline keerukus . . . . .	23
3.2	Kirjeldatud testide võimsuste võrdlemine . . . . .	25
	<b>Kokkuvõte</b>	<b>29</b>
	<b>Kasutatud kirjandus</b>	<b>30</b>
	<b>Lisad</b>	<b>31</b>
	Lisa 1. Jonckheere-Terpstra statistiku väärtuse leidmiseks rakendatud Pythoni kood . . . . .	31
	Lisa 2. Modifitseeritud Jonckheere-Terpstra statistiku väärtuse leidmiseks rakendatud Pythoni kood . . . . .	32
	Lisa 3. Terpstra-Mageli statistiku väärtuse leidmiseks rakendatud Pythoni kood	33
	Lisa 4. Cuzicki statistiku väärtuse leidmiseks rakendatud Pythoni kood . . .	34
	Lisa 5. Shan-Young-Kangi statistiku väärtuse leidmiseks rakendatud Pythoni kood . . . . .	35
	Lisa 6. Statistikatarkvara R simuleerimise koodi näidis . . . . .	36

## Sissejuhatus

Klassikalise statistika meetodid on laialt levinud ning neid on põhjalikult uuritud. Nende keskmes seisab uuritava tunnuse jaotuse pere fikseerimine. Täpse jaotuse leidmiseks piisab seega ainult jaotuse parameetrite hindamisest. Praktikas osutub täpse jaotuse leidmine tihtipeale väga keeruliseks. Sellisel juhul on mõistlik kaaluda mitteparameetrilisi meetodeid. Mitteparameetrilised meetodid ei nõua uuritava tunnuse kohta täpse jaotuse teadmist, vaid lähtuvad oluliselt üldisematest eeldustest. Seetõttu ongi mitteparameetrilised meetodid praktikas laialt kasutatavad.

Selle töö raames on keskendutud rohkem kui kahe grupi järjestuse kontrollimiseks väljamõeldud mitteparameetrilistele testidele. Rohkem kui kahe grupi järjestuse kontrollimiseks määratakse gruppide eeldatav järjekord, võetakse gruppidest valimid ning kasutatakse mõnda antud töös kirjeldatud meetodeid, et veenduda järjekorra kehtivuses. Oluline on märkida, et gruppide eeldatav järjekord tuleb koostada valimi andmeid vaatamata. Enimlevinud mitteparameetriline test, mis gruppide järjekorda kontrollib, on Jonckheere-Terpstra test.

Töö esimeses osas kirjeldatakse käsitletavaid kontrollitavaid hüpoteese, Jonckheere-Terpstra testi, selle statistiku täpset jaotust ja asümptootilist jaotust suurte valimite korral. Lisaks tuuakse statistiku modifikatsioonid juhu jaoks kui valimis esineb võrdseid väärtusi. Töö teises osas tuuakse välja veel neli gruppide järjestuse kontrollimiseks välja töötatud mitteparameetrilised testid. Kolmas osa keskendub eelnevalt kirjeldatud testide võrdlemisele.

# 1 Jonckheere-Terpstra test

Jonckheere-Terpstra test, mis on teadaolevalt üks esimesi teste rohkem kui kahe grupi järjestuse kontrollimiseks, leidis esialgselt kasutust peamiselt meditsiinis ja psühholoogias.[1] Jonckheere-Terpstra testi kasutuselaseks näiteks sobib inimesete operatsioonijärgse taastumise aja järjestuse kontrollimine, sõltuvalt taastumise käigus võetud ravimitest. Võetud ravimite põhjal saab jaotada inimesed erinevatesse rühmadesse, fikseerida rühmade taastumise järjestuse ning järjestust Jonckheere-Terpstra testi abil kontrollida.

Jonckheere-Terpstra teststatistik on tegelikult välja töötatud T. J. Terpstra poolt. Jonckheere'i teststatistik ja Terpstra teststatistik on omavahel seotud lineaarteisenduse kaudu, mis teeb ühest statistikust teise tuletamise väga kergeks.[4] Seetõttu on ka selles peatükis Jonckheere'i ja Terpstra testid eraldi kirjeldatud.

## 1.1 Hüpoteesid ja eeldused

Olgu võrreldud  $k$  erinevat gruppi. Testi statistiku leidmiseks võetud valim koosneb

$$N = \sum_{j=1}^k n_j \quad j = 1, \dots, k$$

elemendist, kus  $n_j$  on  $j$ -nda valimi maht. Olgu  $j$ -nda valimi elemendid  $(x_{1j}, x_{2j}, \dots, x_{n_j})$ .

Testi jaoks tehtavad eeldused on:

1. valimi elemendid  $(x_{1j}, x_{2j}, \dots, x_{n_j})$ , kus  $j = 1, \dots, k$  on sõltumatud,
2. iga fikseeritud  $j = 1, \dots, k$  korral, valimi elemendid  $(x_{1j}, x_{2j}, \dots, x_{n_j})$  on samast pidevast jaotusest, jaotusfunktsiooniga  $F_j$ ,

3. jaotusfunktsioonid  $F_1, \dots, F_k$  on seoses

$$F_j(t) = F(t - \tau_j), \quad -\infty < t < \infty,$$

iga  $j = 1, \dots, k$  korral, kus  $F$  on jaotusfunktsioon pidevale jaotusele tundmatu mediaaniga  $\theta$  ja  $\tau_j$  on  $j$ -nda grupi teadmata mõju jaotusfunktsiooni argumentidele.[2, lk 190]

Jonckheere-Terpstra testi korral on nullhüpoteesiks  $H_0 : [\tau_1 = \dots = \tau_k]$  ehk grupid ei eristu vaadeldava tunnuse poolest. Alternatiivseks hüpoteesiks on  $H_1 : [\tau_1 \leq \tau_2 \leq \dots \leq \tau_k]$ , kus vähemalt üks võrdustest on range.[1] Seega kontrollib Jonckheere-Terpstra test, kas gruppide mõjud tundmatule jaotusfunktsioonile  $F$  on eeldatavas järjekorras. Oluline on märkida, et järjekord  $\tau_1, \dots, \tau_k$  on fikseeritud enne valimi võtmist ja ei ole mõjutatud valimi andmetest.

## 1.2 Jonckheere'i test

Jonckheere'i test baseerub statistikul, mis kasutab Mann-Whitney statistikut summeerides seda üle kõikvõimalike unikaalsete valimi paaride. Mann-Whitney statistik on kujul

$$U_{st} = \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} I(x_{is} < x_{jt}), \quad (1)$$

kus  $x_{is}$  on  $s$ -nda valimi  $i$ -s element ja funktsion  $I(a < b)$  on indikaatorfunktsioon.[3] Esialgu on vaadeldud juhtu, kus valimites kordusi ei esine. Võrdsete elementide juhtumile pööratakse tähelepanu hilisemates peatükkides.

Mann-Whitney statistikut kasutab Mann-Whitney test, mille abil on võimalik kontrollida, kas kaks gruppi erinevad üksteisest või mitte.[3] Seepärast on ka loomulik, et

rohkem kui kahe grupi järjestuse kontrollimiseks kasutatav test baseerub Mann-Whitney testil.

Jonckheere'i pakutud testi statistik on kujul:

$$S = 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij} - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i n_j. \quad (2)$$

Statistiku väärtuse arvutamiseks leitakse vastavad Mann-Whitney statistikud valimite vahel ja lahutatakse igast ühest maha antud valimite mahtude korrutis. Valimi mahtude korrutise lahutamine statistikust teostatakse eesmärgiga võrdsustada statistiku keskväertus nulliga.[1]

## Näide 1

Statistiku väärtuse arvutamise loogikast parema arusaamise tagamiseks on toodud järgmine näide. Olgu järjestatud neli gruppi ja igast grupist olgu võetud viieelemendiline valim. Saadud tulemus olgu järgmine:

I	II	III	IV
9	10	13	16
12	14	15	17
19	21	20	18
28	29	30	27
34	31	40	42

Soovitakse kontrollida hüpoteesi  $H_1 : [\tau_I < \tau_{II} < \tau_{III} < \tau_{IV}]$ . Leiame Jonckheere'i statistiku  $S$  väärtuse. Kasulik on märkida, et  $n_I = n_{II} = n_{III} = n_{IV} = 5$ . Sellisel juhul

$$U_{12} = \sum_{i=1}^5 \sum_{j=1}^5 I(x_{1i} < x_{2j}) = 14$$

ning

$$U_{13} = 16 \quad U_{14} = 14 \quad U_{23} = 14 \quad U_{24} = 14 \quad U_{34} = 14.$$

Seega lähtudes valemist (2) saadakse, et

$$S = 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij} - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i n_j = 2(14 + 16 + 14 + 14 + 14 + 14) - 6 * 25 = 22.$$

### 1.3 Terpstra statistik

T. J. Terpstra [4] defineeris oma statistiku kui

$$T = \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij}.$$

On lihtne näha, et

$$S = 2T - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i n_j. \quad (3)$$

Tänu leitud seosele on järeldatav ka, et

$$ES = 2ET - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i n_j \quad (4)$$

ja

$$DS = 4DT. \quad (5)$$

Seega avaldub Terpstra teststatistik lineaarteisenduse kaudu Jonckheere'i teststatistikust. Järelikult on statistikute täpne ja asümptootiline jaotus eeldatavasti samad.

Selle seose tõttu kutsutakse ka Jonckheere'i ja Terpstra testi kombineeritud nimega Jonckheere-Terpstra test. Rohkem kasutatakse Terpstra statistikut, sest seda on kergem arvutada.[2, lk 202]

#### 1.4 Jonckheere'i ja Terpstra statistikute täpne jaotus nullhüpoteesi kehtides, ilma võrdsete valimi väärtusteta

Jonckheere'i ja Terpstra statistikute täpse jaotuse leidmine osutub valimi mahtude kasvades väga töömahukaks. Näiteks juba kolme grupi korral, kus igast grupist on võetud 5 liiget, on täpse jaotuse leidmiseks tarvis läbi vaadata

$$\frac{N!}{\prod_{i=1}^k n_j!} = \frac{15!}{5!5!5!} = 756756$$

erinevat kombinatsiooni.[2, lk 207]

Kombinatsioonide arv järeldeb asjaolust, et nullhüpoteesi korral on iga  $N$  elemendi ja-gunemise kombinatsioon  $k$  valimisse valimimahtudega  $n_1, n_2, \dots, n_k$  võrdse tõenäosusega. Täpse jaotuse leidmine on läbi tehtud näite korral, kus esimese grupi valim koosneb ühest elemendist, teise grupi valim ühest elemendist ja kolmanda grupi valim kahest elemendist. Kuigi Jonckheere'i ja Terpstra statistikute väärtuse leidmisel ei kasutata valimi elementide astakuid ühisest variatsioonreast, siis täpse jaotuse leidmiseks on mugav astakuid kasutada. Hetkel on võimalik valimi elementide astakutest saada

$$\frac{4!}{1!1!2!} = 12$$

erinevat kombinatsiooni.[2, lk 207]

Tabel 1: Terpstra statistiku väärtused erinevate astakute kombinatsioonide puhul

1)	I II III 1 2 3 4	T=5	2)	I II III 1 3 2 4	T=4
3)	I II III 1 4 2 3	T=3	4)	I II III 2 1 3 4	T=4
5)	I II III 2 3 1 4	T=3	6)	I II III 2 4 1 3	T=2
7)	I II III 3 1 2 4	T=3	8)	I II III 3 2 1 4	T=2
9)	I II III 3 4 1 2	T=1	10)	I II III 4 1 2 3	T=2
11)	I II III 4 2 1 3	T=1	12)	I II III 4 3 1 2	T=0

Tabelis 1 on toodud Terpstra statistiku väärtus erinevate astakute kombinatsioonide jagunemisel valimite I, II ja III vahel. Tabeli 1 järgi saame leida Terpstra statistiku täpse jaotuse, mis on toodud tabelis 2.

Tabel 2: Terpstra statistiku täpne jaotus nullhüpoteesi kehtimise korral

$T_i$	0	1	2	3	4	5
$p_i$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$

Kasutades seost (3) on Terpstra statistiku jaotusest otseselt järeldatav ka Jonckheere'i statistiku täpne jaotus, mis on välja toodud tabelis 3.

Tabel 3: Jonckheere'i statistiku täpne jaotus nullhüpoteesi kehtimise korral

$S_i$	-5	-3	-1	1	3	5
$p_i$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$

Oluline on märgata, et Jonckheere'i ja Terpstra statistikute täpsed jaotused nullhüpoteesi kehtides on samad. Tabelis 3 on näha, et jaotused on sümmeetrilised. See tuleneb asjaolust, et mõlema statistiku arvutamiseks kasutatakse Mann-Whitney statistiku jaotus on samuti sümmeetriline. [3]

## 1.5 Jonckheere'i ja Terpstra statistiku keskväärtus ja dispersioon nullhüpoteesi kehtides

Jonckheere'i ja Terpstra statistik baseerub Mann-Whitney statistikul. Valimeid  $i$  ja  $j$  võrdleva Mann-Whitney statistiku keskväärtus nullhüpoteesi kehtides on H. B. Manni ja D. R. Whitney artikli [3] põhjal

$$EU_{ij} = \frac{n_i n_j}{2}.$$

Seega nullhüpoteesi kehtides saame Terpstra statistiku keskvärtuseks

$$\begin{aligned} ET &= E \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^k EU_{ij} = \sum_{i=1}^{k-1} \sum_{i=j+1}^k \frac{n_i n_j}{2} = \\ &= \frac{1}{4} \left( \sum_{i=1}^k \sum_{j=1}^k n_i n_j - \sum_{i=1}^k n_i^2 \right) = \frac{1}{4} \left( N^2 - \sum_{i=1}^k n_i^2 \right). \end{aligned}$$

Lähtudes seosest (4) saame, et

$$ES = 2ET - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i n_j = 2ET - \frac{1}{2} \left( \sum_{i=1}^k \sum_{j=1}^k n_i n_j - \sum_{i=1}^k n_i^2 \right) = 0.$$

Terpstra statistiku dispersiooni leidmine nullhüpoteesi kehtides nõuab palju algebra-  
list tööd ja seetõttu ei ole antud töös tuletuskäiku välja toodud. Terpstra statistiku  
dispersioon avaldub T. J. Terpstra artikli [4] põhjal kujul:

$$DT = \frac{N^2(2N + 3) - \sum_{j=1}^k n_j^2(2n_j + 3)}{72}.$$

Lähtudes seosest (5) saame, et Jonckheere'i dispersioon avaldub kujul

$$DS = 4DT = \frac{N^2(2N + 3) - \sum_{j=1}^k n_j^2(2n_j + 3)}{18}.$$

## 1.6 Asümptootiline jaotus

Suurte valimi mahtude korral võib Jonckheere'i ja Terpstra statistikute täpse jaotuse  
leidmine kujuneda väga töömahukaks. Sellepärast on mõistlik leida statistikute piir-  
jaotus nullhüpoteesi kehtides.[2, lk 203] Tuleb välja, et kui  $\min(n_1, \dots, n_k) \rightarrow \infty$ , siis

nullhüpoteesi kehtides

$$T^* = \frac{T - ET}{\sqrt{DT}} \xrightarrow{D} N(0, 1).$$

Samuti kui  $\min(n_1, \dots, n_k) \rightarrow \infty$ , siis

$$S^* = \frac{S - ES}{\sqrt{DS}} \xrightarrow{D} N(0, 1).$$

Nullhüpotees kukutatakse, kui  $T^* \geq \bar{z}_\alpha$  või  $S^* \geq \bar{z}_\alpha$ , kus  $\alpha$  on olulisuse nivoo ja  $\bar{z}_\alpha$  on  $N(0, 1)$  täiendkvantiil.[2, lk 203]

## Näide 2

Järgnevalt on näite 1 andmete põhjal leitud Terpstra statistiku asümptootiline jaotus. Näites võeti võrdsete suurustega viieliikmeline valim kõigist neljast grupist. Täpse jaotuse leidmiseks tuleks läbivaadata

$$\frac{20!}{5!5!5!5!} \approx 1,2 \times 10^{10}$$

erinevat kombinatsiooni. Selle asemel on mõistlik kasutada asümptootilist jaotust. Näites 1 on leitud Jonckheere'i statistiku väärtus. Seose (3) põhjal  $T = 86$ . Lisaks on vaja veel leida  $ET$  ja  $DT$ . Lähtudes eelmise peatüki valemitest on saadud, et

$$ET = \frac{1}{4}(20^2 - 4 \times 5^2) = 75$$

ja

$$DT = \frac{20^2(2 \times 20 + 3) - 4 \times 5^2(2 \times 5 + 3)}{72} \approx 220,8.$$

Seega  $T^* = \frac{86-75}{\sqrt{220,8}} \approx 0,74$ . Olulisuse nivool 0,05 on  $\bar{z}_{0,05} = 1,65$ . Kuna  $T^* < \bar{z}_{0,05}$ , siis antud näites nullhüpoteesi ei kukutata.

## 1.7 Võrdsete väärtused

Siiani on kõik töös toodud tulemused saadud eeldusel, et valimites ei leidu võrdseid väärtusi. Selline kitsendus ei ole kindlasti praktiline ja hea oleks saadud tulemused üldistada ka juhule kui valimites leidub võrdseid väärtusi. Selleks saab indikaatorfunktsioon valemis (1) võrdsete argumentide korral väärtuseks  $\frac{1}{2}$ .

Korduvate väärtuste esinemisel muutub Terpstra statistiku dispersioon, mis avaldub kujul

$$\begin{aligned} DT &= \frac{1}{72} \left( N(N-1)(2N+5) - \sum_{j=1}^k n_j(n_j-1)(2n_j+5) - \sum_{h=1}^g t_h(t_h-1)(2t_h+5) \right) + \\ &+ \frac{1}{36N(N-1)(N-2)} \left( \sum_{j=1}^k n_j(n_j-1)(n_j-2) \right) \left( \sum_{h=1}^g t_h(t_h-1)(t_h-2) \right) + \quad (6) \\ &+ \frac{1}{8N(N-1)} \left( \sum_{j=1}^k n_j(n_j-1) \right) \left( \sum_{h=1}^g t_h(t_h-1) \right), \end{aligned}$$

kus  $g$  näitab võrdsete väärtuste gruppide arvu ja  $t_h$  on  $h$ -nda võrdsete väärtuste grupi suurus.[2, lk 204] Jonckheere'i statistiku dispersioon korduvate väärtuste esinemisel avaldub seosest (5).

### Näide 3

Järgnevalt on toodud dispersiooni valemi (6) paremaks arusaamiseks lühike näide. Olgu kolm gruppi, kusjuures igast ühest on võetud neljaelemendiline valim. Saadud valimid on

I	II	III
2	3	4
5	5	5
6	7	8

Jonckheere'i ja Terpstra statistiku leidmiseks on ära märgitud, et  $n_1 = n_2 = n_3 = 3$  ning  $U_{12} = 5.5$ ,  $U_{13} = 5.5$  ja  $U_{23} = 5.5$ . Terpstra statistiku väärtuseks on

$$T = \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij} = 5.5 + 5.5 + 5.5 = 16.5.$$

Ka võrdsete valimite korral kehtib seos (3) ehk

$$S = 2T - 3 \times 9 = 6$$

Kuna statistiku keskväärtus võrdsetest väärtustest ei olene, siis

$$ET = \frac{1}{4}(9^2 - 3 \times 9) = 13,5.$$

Seose (6) põhjal on leitud Terpstra statistiku dispersioon:

$$\begin{aligned} DT &= \frac{1}{72}[9(9-1)(18+5) - 3 \times 3(3-1)(6+5) - 3(3-1)(6+5)] + \\ &+ \frac{1}{36 \times 9(9-1)(9-2)}[3 \times 3(3-1)(3-2)][3(3-1)(3-2)] + \\ &+ \frac{1}{8 \times 9(9-1)}[3 \times 3(3-1)][3(3-1)] \approx 19,53. \end{aligned}$$

Võrdluseks on leitud sama näite põhjal ka Terpstra statistiku dispersioon juhul, kui korduvaid väärtusi ei eksisteeriks:

$$DT_* = \frac{N^2(2N+3) - \sum_{j=1}^k n_j^2(2n_j+3)}{72} = \frac{9^2(18+3) - 3 \times 9(6+3)}{72} = 20,25.$$

Seega on näha, et korduvate väärtuste esinemisel statistiku dispersioon väheneb.

## 2 Veel teste rohkem kui kahe grupi järjestuse kontrollimiseks

Selles peatükis on lühidalt kirjeldatud veel nelja grupi järjestuse kontrollimiseks välja töötatud testi. Iga kirjeldatud testi puhul on läbimängitud ka statistiku arvutamine näites 1 esitatud andmete põhjal.

### 2.1 Modifitseeritud Jonckheere-Terpstra test

Modifitseeritud Jonckheere-Terpstra test baseerub Jonckheere-Terpstra testil aga arvestab ka grupid omavahelist kaugust. Omavahelise kauguse all on mõeldud gruppide kaugust üksteisest testitavas järjekorras. Kauguse arvestamise mõte tuleneb asjaolust, et grupid on järjestatud kontrollitavas järjekorras. Seega omavahel kaugemate gruppide Mann-Whitney statistiku väärtus saab suurema kaalu, sest on olulisem kui naabergruppide Mann-Whitney statistiku väärtus. M. Neuhäuser näitas, et väikeste valimite korral on modifitseeritud Jonckheere-Terpstra test võimsam kui harilik Jonckheere-Terpstra test. Valimi mahtude kasvades, võimsuse erinevus kaob.[5]

Modifitseeritud Jonckheere-Terpstra statistik avaldub kujul

$$MJT = \sum_{i=1}^{k-1} \sum_{j=i+1}^k (j-i)U_{ij},$$

kus  $U_{ij}$  on  $i$ -nda ja  $j$ -nda grupist võetud valimite võrdlemise Mann-Whitney statistik.[5]

## Näide 4

Järgmisena on näite 1 põhjal leitud modifitseeritud Jonckheere-Terpstra statistiku väärtus. Näites 1 on välja toodud ka, et

$$U_{12} = 14 \quad U_{13} = 16 \quad U_{14} = 14 \quad U_{23} = 14 \quad U_{24} = 14 \quad U_{34} = 14.$$

Seega

$$MJT = (2 - 1) \times U_{12} + (3 - 1) \times U_{13} + (4 - 1) \times U_{14} + (3 - 2) \times U_{23} + (4 - 2) \times U_{24} + (4 - 3) \times U_{34} = 144.$$

Modifitseeritud Jonckheere-Terpstra statistikut on võimalik leida ka võrdsete valimite väärtuste korral. Lähenemine on analoogne Jonckheere-Terpstra statistiku leidmisele võrdsete väärtuste korral.[5]

## 2.2 Terpstra-Mageli test

Erinevalt Jonckheere-Terpstra testist, kus vaadatakse üksahaaval läbi kõik valimite paarid, on Terpstra-Mageli test üles ehitatud gruppidest võetud valimite samaaegsele võrdlemisele. Terpstra-Mageli statistik avaldub kujul:

$$TM = \sum_{m_1=1}^{n_1} \sum_{m_2=1}^{n_2} \cdots \sum_{m_k=1}^{n_k} I(x_{1m_1} \leq x_{2m_2} \leq \cdots \leq x_{km_k}),$$

kus  $I(a \leq \cdots \leq b)$  tähistab indikaatorfunktsiooni.[6]

Terpstra-Mageli testi suurimaks puuduseks on statistiku arvutamiseks kuluv töömaht. Statistiku arvutamiseks leitavas summas on kokku  $n_1 n_2 \dots n_k$  liidetavat. Igas liidetavas

tehakse  $k - 1$  võrdlemist.[6] Näite 1 põhjal Terpstra-Mageli statistiku väärtuse leidmine osutub seetõttu keeruliseks, sest vajab  $5^4 \times 3 = 1875$  võrdlemist. Hariliku Jonckheere-Terpstra statistiku väärtuse leidmiseks tehakse  $6 \times 5^2 = 150$  võrdlemist. Seega võib öelda, et Terpstra-Mageli statistiku väärtuse leidmine on oluliselt ajamahukam, kui hariliku Jonckheere-Terpstra statistiku väärtuse leidmine.

Terpstra-Mageli statistiku ülesehitusest on näha, et statistik arvestab ka võrdsete väärtustega valimites.

### 2.3 Cuzicki test

Cuzicki test lähtub Wilcoxon'i astaksumma testist ja on kolmas erinev lähenemine gruppide järjestuse kontrollimiseks. Cuzicki teststatistik avaldub kujul

$$CU = \sum_{i=1}^k i \sum_{j=1}^{n_j} R_{ij},$$

kus  $R_{ij}$  on  $i$ -nda grupi  $j$ -nda elemendi astak üldises variatsioonreas.[6]

### Näide 5

Järgnevalt on leitud näite 1 põhjal Cuzicki statistiku väärtus. Statistiku väärtuse leidmiseks on esmalt koostatud variatsioonrida:

$$\begin{array}{cccccccccc} 9 & 10 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 \\ 20 & 21 & 27 & 28 & 29 & 30 & 31 & 34 & 40 & 42 \end{array} \quad (7)$$

Statisiku väärtuseks on

$$CU = 1 \times (3 + 10 + 1 + 14 + 18) + 2 \times (5 + 12 + 2 + 17 + 15) + \\ + 3 \times (4 + 11 + 6 + 16 + 19) + 4 \times (8 + 9 + 7 + 20 + 13) = 544.$$

## 2.4 Shan-Young-Kangi test

Shan-Young-Kangi nimetust kannab test, mille pakkusid välja Guegon Shan, Daniel Young ja Le Kang 2014 aastal.[6] Sellel testil ametlikult muud nime ei eksisteeri. Shan-Young-Kangi test on väga sarnane Jonckheere-Terpstra testile. Kahe testi erinevus seisneb selles, et Shan-Young-Kangi test kaasab statistiku arvutamisesse ka valimi elementide astakud üldises variatsioonreas, ning pöörab ümber indikaatorfunktsiooni otuse. Shan-Young-Kangi teststatistik avaldub kujul

$$K = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} (R_{jm} - R_{il}) I(x_{jm} > x_{il}),$$

kus  $R_{jm}$  on  $j$ -nda grupi valimi  $m$ -nda elemendi astak ja  $I(a > b)$  on indikaatorfunktsioon.[6]

## 3 Kirjeldatud testide võrdlemine

### 3.1 Kirjeldatud testide ajaline keerukus

Mitte vähetähtsam on suuremahuliste andmete analüüsimisel analüüsimeetodi keerukus. Meetodi keerukusest saame rääkida kui leidub algoritm, mida antud meetod rakendab. Lisades 1 kuni 5 on välja toodud Pythoni programmeerimiskeeles kirjutatud algoritmid eelnevalt kirjeldatud testide statistikute väärtuse leidmiseks.

Algoritmi ajalist keerukust hinnatakse funktsiooniga  $c(n)$ , kus  $n$  on töödeldavate andmete maht (vaadeldaval juhul  $N$ , mis on valimite mahtude summa). Ajalise keerukuse hindamiseks leitakse, millise funktsiooniga on  $c(n)$  alt ja ülalt asümptootiliselt tõkestatud. Asümptootilist alt ja ülalt tõkestatust tähistatakse  $c(n) \in \Theta(f(n))$ , mis tähendab, et funktsioon  $c(n)$  on asümptootiliselt alt ja ülalt tõkestatud funktsiooniga  $f(n)$ . [7, lk 21]

Ajalise keerukuse leidmise lihtsustamiseks on eeldatud, et kontrollitakse  $k$  erineva grupi järjekorda, ning et igast grupist võetud valimi maht on  $n$ .

Jonckheere-Terpstra statistiku leidmiseks rakendatav Pythoni programmeerimiskeele algoritm on leitav lisast 1. Jonckheer-Terpstra statistiku leidmise algoritmi ajaline keerukus on  $c(n) \in \Theta(k^2n^2)$ , sest statistiku leidmiseks läbitakse  $\frac{k(k-1)}{2}$  korda funktsiooni  $MW$ , mille ajalise keerukuse funktsioon on  $c(n) \in \Theta(n^2)$ . Funktsiooni  $MW$  ajaline keerukus tuleneb topelt tsüklilise funktsiooni sees, mille mõlema pikkus on  $n$  iteratsiooni.

Modifitseeritud Jonckheere-Terpstra statistiku leidmiseks rakendatav Pythoni programmeerimiskeele algoritm on leitav lisast 2. Modifitseeritud Jonckheere-Terpstra statistiku

leidmise algoritmi ajaline keerukus on võrdne Jonckheere-Terpstra statistiku leidmise algoritmi ajalise keerukusega, sest algoritmid erinevad ainult elementaartehete poolest.

Terpstra-Mageli statistiku algoritmi, mis on leitav lisast 3, ajaline keerukus on  $c(n) \in \Theta(n^k)$ . Ajaline keerukus tuleneb asjaolust, et statistiku leidmiseks läbitakse eelmise tsükli sees tsükli, mille pikkuseks on  $n$  iteratsiooni,  $k$  korda.

Cuzicki statistiku algoritm, mis on kirjeldatud lisas 4, keerukus on  $c(n) \in \Theta(k^2 n^2)$ , sest statistiku leidmiseks läbitakse  $k$  tsükli pikkusega  $n$  iteratsiooni ning igas tsükli iteratsioonis tehakse elemendi astaku otsimist variatsioonreast, mille kõige optimaalsemaks keerukuseks on  $c(n) \in \Theta(N)$ , kus  $N = kn$ . [8]

Shan-Young-Kangi statistiku leidmise algoritmi ajaline keerukus on  $c(n) \in \Theta(k^3 n^3)$ . Algoritm on leitav lisast 5 ja tema keerukus seisneb  $\frac{k(k-1)}{2}$  korda funktsiooni  $RMW$  rakendamisel. Funktsioon  $RMW$  keerukus on  $c(n) \in \Theta(kn^3)$ , sest funktsiooni sees läbitakse topelt tsükli pikkusega  $n$  iteratsiooni ja igas tsükli iteratsioonis leitakse variatsioonreast element, mille kõige optimaalsemaks keerukuseks on  $c(n) \in \Theta(N)$ , kus  $N = kn$ . [8]

Praktikas on ratsionaalne arvata, et gruppide arv  $k$  on üldjuhul väike. Suureks võivad minna valimi mahud. Selle põhjal tuleks eelistada kindlasti Jonckheere-Terpstra, modifitseeritud Jonckheere-Terpstra ning Cuzicki teste, sest nende ajaline keerukus on väikse  $k$  ja suure  $n$ -i korral kõige optimaalsem. Keerukuselt järgneb Shan-Young-Kangi test ja kõige keerukamaks osutus Terpstra-Mageli test.

## 3.2 Kirjeldatud testide võimsuste võrdlemine

Kirjeldatud testide võimsuste võrdlemiseks on erinevatest teadaolevatest jaotustest simuleeritud valimid ja leitud kui suure tõenäosusega mingi test valimite järjestuse, mis tuleneb jaotuse mediaanide järjestusest, õigeks loeb. Võimsuse leidmiseks on valimi võtmist korratud 10000 korda. Olulisuse nivooks on võetud 0,05 ja testide otsused on leitud statistikatarkvara R poolt väljastatud p-väärtuste põhjal. Andmed on genereeritud nii nullhüpoteesile vastavast olukorrast kui ka erinevatest alternatiivse hüpoteesi variantidest.

Simuleerimine on läbiviidud statistikatarkvaraga R ja näide simuleerimise koodst on toodud lisas 6. Kuna Terpstra-Mageli testile ja Shan-Young-Kangi poolt välja pakutud testile vastavat funktsiooni statistikatarkvaras R ei eksisteeri, siis on Terpstra-Mageli testi ja Shan-Young-Kangi testi võimsused võetud G. Shani, D. Youngi ja L. Kangi artiklist [6], kus on lisaks leitud ka teiste antud töös kirjeldatud testide võimsused. Kuna simuleeritud võimsused artiklis [6] on väga sarnased antud töö raames simuleeritud võimsustele, on alust arvata, et edasised võrdlused on usaldusväärsed.

Esimesed simulatsioonid on läbiviidud kolme normaaljaotuse põhjal. Simuleerimistel kasutatud normaaljaotuste dispersioon on 1. Dispersioon erineb ainult viimasel juhul, kus andmed on simuleeritud kolmest, nii keskvaärtuse kui ka dispersiooni poolest, erinevast normaaljaotusest. Täielikud simuleerimise tulemused on välja toodud tabelis 4.

Tabel 4: Kirjeldatud testide võimsused normaaljaotusest simuleeritud valimite korral

Jaotus	Valimi mahud	JT	MJT	TM	CU	S
$\mu=(0,0,0)$	(10,10,10)	0.052	<b>0.053</b>	0.049	0.046	0.051
	(10,15,20)	<b>0.048</b>	<b>0.048</b>	<b>0.048</b>	0.047	<b>0.048</b>
$\mu=(0,0.5,1)$	(10,10,10)	0.661	0.664	0.601	0.662	<b>0.679</b>
	(10,15,20)	0.790	0.795	0.723	0.800	<b>0.804</b>
$\mu=(0,1,1)$	(10,10,10)	0.637	0.650	0.527	0.647	<b>0.684</b>
	(10,15,20)	0.645	<b>0.674</b>	0.649	0.664	0.653
$\mu=(0,0,1)$	(10,10,10)	0.632	0.643	0.519	0.646	<b>0.683</b>
	(10,15,20)	0.885	0.886	0.625	0.865	<b>0.902</b>
$\mu=(0,1,0.5)$	(10,10,10)	0.224	0.230	0.144	0.232	<b>0.290</b>
	(10,15,20)	0.110	0.155	0.152	<b>0.163</b>	0.143
N(0,9);N(0.6,4);N(1,1)	(10,10,10)	0.133	0.125	<b>0.186</b>	0.119	0.153
	(10,15,20)	0.164	0.164	0.229	0.164	<b>0.244</b>

Simuleerimiste tulemustena on Shan-Young-Kangi test osutunud kõige võimsamaks kaheksal juhul kaheteistkümnest. Kuuel korral kaheteistkümnest oli kõige kehvema võimsusega Terpstra-Mageli test. Terpstra-Mageli test on  $\mu=(0,0,1)$  ja valimi mahtude (10,15,20) korral 0,277 võrra väiksema võimsusega kui selle kombinatsiooni parima testi Shan-Young-Kangi võimsus. Jättes välja Terpstra-Mageli testi, varieeruvad kõikide ülejäänud testide võimsused üsna vähe ning ei erine üksteisest rohkem kui 0,1 võrra. Nii väike võimsuste varieerumine teeb statistikute leidmise algoritmi keerukuse väärtuslikumaks, sest kui võimsuses palju ei kaotata, on praktikas parem kasutada kõige optimaalsemat meetodit.

Tabelis 5 on toodud järgmiste simuleerimiste tulemused, kus grupid on kõik ekspo-

nentjaotusest. Kõikide statistikute võimsused eksponentjaotuse korral on madalad. Kõige suurema võimsuse saavutas Shan-Young-Kangi test, mille maksimaalne võimsus oli 0,552. Shan-Young-Kangi test osutus ka kõige paremaks, olles kõige võimsam viiel korral kümnest. Kõige kehvemaks osutus taaskord Terpstra-Mageli test, mis oli kõige kehvema võimsusega seitsmel korral kümnest. Jonckheere-Terpstra test oli kõige kehvem neljal korral kümnest. Jonckheere-Terpstra testi kõige suurem võimsuse erinevus võrreldes kõige võimsama testiga oli kõigest 0,035.

Eksponentjaotuse korral on oluline märkida, et jaotuse keskväärtus ei ole nihketa hinnang jaotuse mediaanile. Sellegipoolest on eksponentjaotuse mediaanid samas suuruse järjekorras, mis keskväärtused.

Tabel 5: Kirjeldatud testide võimsused eksponentjaotusest simuleeritud valimite korral

Jaotus	Valimi mahud	JT	MJT	TM	CU	S
$\mu=(1,1,1)$	(10,10,10)	0.046	0.046	<b>0.052</b>	0.047	0.051
	(10,15,20)	0.048	0.048	<b>0.052</b>	0.048	0.051
$\mu=(1,1.5,2)$	(10,10,10)	0.350	0.351	0.321	0.349	<b>0.367</b>
	(10,15,20)	0.428	<b>0.438</b>	0.405	<b>0.438</b>	0.426
$\mu=(1,2,2)$	(10,10,10)	0.326	0.329	0.262	0.329	<b>0.351</b>
	(10,15,20)	0.327	<b>0.362</b>	0.321	0.361	0.335
$\mu=(1,1,2)$	(10,10,10)	0.340	0.343	0.323	0.340	<b>0.362</b>
	(10,15,20)	0.534	0.522	0.409	0.510	<b>0.552</b>
$\mu=(1,2,1.5)$	(10,10,10)	0.150	0.155	0.104	0.157	<b>0.180</b>
	(10,15,20)	0.108	0.130	0.121	<b>0.139</b>	0.121

Järgmistes simuleerimistes on esimese grupi jaotuseks normaaljaotus, teise ja kolmanda

grupi jaotuseks aga eksponentjaotus. Taaskord on testide võimsused tagasihoidlikud ja kõige suurema võimsuse 0,463 saavutab taaskord Shan-Young-Kangi test. Shan-Young-Kangi test on viiel korral kümnest kõige võimsam. Terpstra-Mageli test on kuuel korral kümnest kõige vähem võimasam ning kahel korral kõige võimsam test. Jonckheere-Terpstra test on kolmel korral kümnest kõige vähem võimsam test. Jättes Terpstra-Mageli testi võrdlemisest välja, on võimsuste erinevused taaskord väga väikesed. Teiste testide võimsused erinevad vaadeldud simulatsioonides kõige rohkem 0,045 võrra. Täpsemad tulemused on välja toodud tabelis 6.

Tabel 6: Kirjeldatud testide võimsused segajaotusest simuleeritud valimite korral

Jaotus	Valimi mahud	JT	MJT	TM	CU	S
$\mu=(1,1,1)$	(10,10,10)	0.030	0.029	<b>0.045</b>	0.028	0.030
	(10,15,20)	0.027	0.027	<b>0.048</b>	0.029	0.030
$\mu=(1,1.5,2)$	(10,10,10)	0.254	0.250	0.240	0.248	<b>0.265</b>
	(10,15,20)	0.327	<b>0.334</b>	0.300	0.331	0.328
$\mu=(1,2,2)$	(10,10,10)	0.241	0.249	0.200	0.245	<b>0.266</b>
	(10,15,20)	0.241	<b>0.269</b>	0.232	0.265	0.243
$\mu=(1,1,2)$	(10,10,10)	0.251	0.256	0.267	0.256	<b>0.281</b>
	(10,15,20)	0.441	0.418	0.327	0.418	<b>0.463</b>
$\mu=(1,2,1.5)$	(10,10,10)	0.099	0.103	0.074	0.102	<b>0.118</b>
	(10,15,20)	0.068	0.083	0.078	<b>0.084</b>	0.076

## Kokkuvõte

Jonckheere-Terpstra testi puhul on antud töös leitud tema statistiku täpne ja asümptootiline jaotus. Valimi mahtude kasvades ei ole tarbekas täpse jaotuse leidmine ja otsuste tegemisel tuleks lähtuda asümptootilisest jaotusest. Asümptootiline jaotus on leitud ka juhul, kui valimites esineb võrdseid väärtusi.

Simulatsioonide põhjal osutub kõige vähem võimsaks testiks Terpstra-Mageli test. Lisaks on Terpstra-Mageli statistiku arvutamise algoritm ajaliselt kõige keerukam. Kuna ülejäänud testide võimsused on väga väikeste erinevustega, siis on pööratud tähelepanu hoopis testide statistikute arvutamise algoritmide keerukusele. Vaadates algoritmide ajalisi keerukusi võib öelda, et Shan-Young-Kangi teststatistiku leidmine on alati ajaliselt keerukam kui Jonckheere-Terpstra, modifitseeritud Jonckheere-Terpstra ja Cuzicki teststatistiku leidmine. Seega tuleks suurte valimi mahtude korral Shan-Young-Kangi testile eelistada hoopis Jonckheere-Terpstra, modifitseeritud Jonckheere-Terpstra või Cuzicki testi.

## Kasutatud kirjandus

1. Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika* 41: 133-145.
2. Hollander, M., Wolfe, D. A. (1999). *Nonparametric Statistical Methods* (2. väljanne). John Wiley & Sons, Inc.
3. Mann, H. B., Whitney, D. R. (1947). On a test of whether one of 2 random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18: 50-60.
4. Terpstra, T. J. (1952) The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one raking. *Indagationes Mathematicae* 14: 327-333.
5. Neuhäuser, M., Liu P. Y. (1998). Nonparametric tests for trend: Joncheere's test, a modification and maximum test. *Biometrical Jorunal* 40: 899-909.
6. Shan, G., Young, D., Kang, L. (2014) A new powerful nonparametric rank test for ordered alternative problem.  
<https://pdfs.semanticscholar.org/bf75/135fce5cf7097f0b78239789fc32a1e9de4e.pdf>  
(7.05.2018).
7. Sander, P., Mehlhorn, K. (2008). *Data Structures and Algorithms*. Springer.
8. Big-O Cheet Sheet <http://bigocheatsheet.com/> (7.05.2018)

## Lisad

### Lisa 1. Jonckheere-Terpstra statistiku väärtuse leidmiseks rakendatud Pythoni kood

```
def MW (valim1 , valim2):  
    MWvalue = 0  
    for element1 in valim1:  
        for element2 in valim2:  
            if element1 < element2:  
                MWvalue += 1  
    return MWvalue
```

```
JT = 0  
valim1 = [9,12,19,28,34]  
valim2 = [10,14,21,29,31]  
valim3 = [13,15,20,30,40]  
valim4 = [16,17,18,27,42]  
JT += MW(valim1 , valim2)  
JT += MW(valim1 , valim3)  
JT += MW(valim1 , valim4)  
JT += MW(valim2 , valim3)  
JT += MW(valim2 , valim4)  
JT += MW(valim3 , valim4)  
print (JT)
```

## Lisa 2. Modifitseeritud Jonckheere-Terpstra statistiku väärtuse leidmiseks rakendatud Pythoni kood

```
def MW(valim1, valim2):  
    MW = 0  
    for element1 in valim1:  
        for element2 in valim2:  
            if element1 < element2:  
                MW += 1  
    return MW
```

```
MJT = 0  
valim1 = [9,12,19,28,34]  
valim2 = [10,14,21,29,31]  
valim3 = [13,15,20,30,40]  
valim4 = [16,17,18,27,42]  
MJT += MW(valim1, valim2)  
MJT += 2*MW(valim1, valim3)  
MJT += 3*MW(valim1, valim4)  
MJT += MW(valim2, valim3)  
MJT += 2*MW(valim2, valim4)  
MJT += MW(valim3, valim4)  
print(MJT)
```

### Lisa 3. Terpstra-Mageli statistiku väärtuse leidmiseks rakendatud Pythoni kood

```
TM= 0

valim1 = [9,12,19,28,34]
valim2 = [10,14,21,29,31]
valim3 = [13,15,20,30,40]
valim4 = [16,17,18,27,42]

for element1 in valim1:
    for element2 in valim2:
        for element3 in valim3:
            for element4 in valim4:
                times = times +3
                if element1 <= element2 and element2 <= element3
                and element3 <= element4:
                    TM += 1

print (TM)
```

## Lisa 4. Cuzicki statistiku väärtuse leidmiseks rakendatud Pythoni kood

```
CU = 0
valim1 = [9,12,19,28,34]
valim2 = [10,14,21,29,31]
valim3 = [13,15,20,30,40]
valim4 = [16,17,18,27,42]

rank = valim1+valim2+valim3+valim4

rank.sort()

for element1 in valim1:
    CU += rank.index(element1)+1

for element2 in valim2:
    CU += 2*(rank.index(element2)+1)

for element3 in valim3:
    CU += 3*(rank.index(element3)+1)

for element4 in valim4:
    CU += 4*(rank.index(element4)+1)

print(CU)
```

## Lisa 5. Shan-Young-Kangi statistiku väärtuse leidmiseks rakendatud Pythoni kood

```
def RMW(valim1, valim2, rank):
    MW = 0
    for element1 in valim1:
        for element2 in valim2:
            if element1 > element2:
                MW += rank.index(element1) - rank.index(element2)
    return MW

S = 0
valim1 = [9,12,19,28,34]
valim2 = [10,14,21,29,31]
valim3 = [13,15,20,30,40]
valim4 = [16,17,18,27,42]
rank = valim1+valim2+valim3+valim4
rank.sort()

S += RMW(valim1, valim2, rank)
S += RMW(valim1, valim3, rank)
S += RMW(valim1, valim4, rank)
S += RMW(valim2, valim3, rank)
S += RMW(valim2, valim4, rank)
S += RMW(valim3, valim4, rank)
print(S)
```

## Lisa 6. Statistikatarkvara R simualeerimise koodi näidis

```
library("clinfun")
library("gMWT")
library("PMCMRplus")

grupp1 <- rep(1,10)
grupp2 <- rep(2,10)
grupp3 <- rep(3,10)
grupp <- c(grupp1, grupp2, grupp3)
power <- rep(0,3)
i <- 0
while (i < 10000) {
  valim1<- rnorm(10,0,9)
  valim2 <- rnorm(10,0.6,4)
  valim3 <- rnorm(10,1,1)
  valim <- c(valim1, valim2, valim3)
  JT <- jonckheere.test(valim,grupp,"increasing")
  MJT <- gmw(valim,grupp, test='jt*', alternative = "greater")
  CU <- cuzickTest(valim, grupp)
  if(JT$p.value < 0.05) {
    power[1] <- power[1] +1
  }
  if(MJT$p.values < 0.05) {
    power[2] <- power[2] +1
  }
}
```

```
if (CU$p.value < 0.05) {  
  power[3] <- power[3] +1  
}  
i <- i+1  
}  
power/10000
```

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

Mina, Kristjan-Martin Kirjanen,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Mitteparameetrilised testid rohkem kui kahe grupi järjestuse kontrollimiseks", mille juhendaja on Anne Selart,
  - (a) reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - (b) üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus 8.05.2018