

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Raul Niit

**Mitteparameetrilised testid kuhja
alternatiivhüpoteesidele**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendaja: Anne Selart

TARTU 2021

MITTEPARAMEETRILISED TESTID KUHJA ALTERNATIIVHÜPOTEESIDELE

Bakalaureusetöö

Raul Niit

Lühikokkuvõte

Käesoleva bakalaureusetöö eesmärk on kirjeldada peamiseid mitteparameetrilisi teste kuhja alternatiivhüpoteeside jaoks ning välja toodud teste omavahel võrrelda. Töö teoreetilises osas esitatakse ülevaade kolmest erinevast teststatistikust ja nende omadustest. Lisaks on toodud näited teststatistikute arvutamise ja täpse jaotuse leidmise kohta. Töö praktilises osas võrreldakse testide käitumist simulatsioonide abil, uurides selleks nii testide võimsust kui ka I liiki vea tegemise tõenäosust erinevate parameetrite korral.

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

Märksõnad: mitteparameetriline statistika, hüpoteeside testimine, simulatsioonid

NON-PARAMETRIC TESTS FOR UMBRELLA ALTERNATIVES

Bachelor thesis

Raul Niit

Abstract

The aim of this bachelor thesis is to describe the main non-parametric tests for umbrella alternatives and compare these tests with each other. In the theoretical part of the thesis, an overview of three different test statistics and their properties are presented. In addition, examples are provided for calculating test statistics and finding the exact distribution. In the practical

part of the thesis, the behavior of these tests are compared using simulations, studying both the power and type I error for different parameters.

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

Key Words: non-parametric statistics, hypothesis testing, simulations

Sisukord

Sissejuhatus	5
1 Teststatistikud kuhja alternatiivhüpoteesi jaoks	7
1.1 Mack-Wolfe test	8
1.1.1 Teststatistik	8
1.1.2 Statistiku täpne jaotus nullhüpoteesi kehtides	10
1.1.3 Statistiku keskvaartus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides	13
1.1.4 Kuhja tipu leidmine	15
1.1.5 Võrdsed väärtused	16
1.2 Hettmansperger-Nortoni test	18
1.2.1 Teststatistik	18
1.2.2 Statistiku täpne jaotus nullhüpoteesi kehtides	20
1.2.3 Statistiku keskvaartus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides	22
1.2.4 Kuhja tipu leidmine	24
1.3 Modifitseeritud Mack-Wolfe statistik	24
1.3.1 Teststatistik	25
1.3.2 Statistiku täpne jaotus nullhüpoteesi kehtides	25
1.3.3 Statistiku keskvaartus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides	27
1.3.4 Kuhja tipu leidmine	28

2	Testide võrdlemine	30
2.1	Andmete genereerimine	30
2.2	Võimsuste võrdlemine	32
2.2.1	Normaaljaotusega andmed	33
2.2.2	EkspONENTjaotusega andmed nihutatud konstandi võrra . . .	38
2.2.3	EkspONENTjaotusega andmed erineva jaotuse parameetri korral	43
2.3	I liiki vea võrdlemine	48
2.3.1	Normaaljaotusega andmed	48
2.3.2	EkspONENTjaotusega andmed	51
	Kokkuvõte	53
	Kasutatud allikad	55
	Lisad	56
A.1	Täpse jaotuse arvutamine	56
A.2	Mack-Wolfe, Hettmannsperger-Nortoni ja modifitseeritud Mack-Wolfe täpne jaotus nullhüpoteesi kehtides näiteandmestiku korral	59
A.3	R kood andmete genereerimiseks	62
A.4	R kood statistikute väärtuste ja testide p-vääruste leidmiseks . . .	64
A.5	R kood testide tegemiseks	69
A.6	R kood jooniste väljastamiseks	70

Sissejuhatus

Paljud statistikas kasutatavad meetodid põhinevad eeldusel, et andmed pärinevad teatud kindlast jaotusest. Praktikas esineb aga tihti olukordi, kus andmete jaotuse kohta eelenev informatsioon puudub. Sellises olukorras võetakse kasutusele mitteparameetiline statistika, kus saab konkreetsest jaotusest pärinemise eeldust lõdvendada ja teha üldisema eelduse, et andmed on mingist pidevast jaotusest.

Käesolevas töös on käsitletud mitteparameetrilisi teste, mis on välja töötatud mitme sõltumatu grupi võrdlemiseks, kui alternatiivne hüpotees väidab seda, et järjestatud grupid paigutuvad uuritava tunnuse keskmise taseme järgi kuhja kujuliselt. Teisisõnu testitakse seda, kas grupid võivad olla järjestatud nii, et kuni mingi teatud grupini p toimub gruppide vahel uuritava tunnuse kasvamine ning alates grupist p hakkab uuritav tunnus gruppide vahel kahanema. Selline olukord võib kehtida näiteks siis, kui inimesed jaotada gruppidesse vanuse järgi ning mõõta nende reaktsioonikiirust - kuni teatud vanuseni lähevad inimese reaktsioonid paremaks, kuid hilisemas elus jääb inimene taas aeglasemaks.

Töö jaguneb kaheks osaks. Töö teoreetilises osas esitatakse Mack-Wolfe, Hettmannsperger -Nortoni ja modifitseeritud Mack-Wolfe testid, mis on mõeldud kuhja alternatiivhüpoteesi testimiseks. Iga testi jaoks tuuakse välja teststatistik, esitatakse statistiku keskväärts ja dispersioon nullhüpoteesi kehtides, uuritakse statistiku täpset ja asümptootilist jaotust ning kirjeldatakse olukorda, kus kuhja tipp p eelnevalt teada ei ole. Kõigi kolme testi juurde on toodud ka näited teststatistiku arvutuskäigu ja täpse jaotuse selgitamiseks.

Töö praktilises osas võrreldakse simulatsioonide abil nimetatud testide võimsusi ning I liiki vea tegemise tõenäosust. Analüüsimisel vaadeldakse variante, kus andmed on tekitatud kolmel erineval viisil, millest üks vastab ka olukorrale, kui testide eeldused on rikutud. Testide käitumist uuritakse varieerides gruppide arvu k , kuhja tipu asukohta p ning vaatluste arvu igas grupis n . Simulatsioonide läbi viimiseks on kasutatud statistikatarkvara R (R Core Team, [2021](#)).

Töös on 6 lisa, kus on esitatud täpse jaotuse arvutamiseks vajalikud tabelid ning analüüsimisel kasutatud tarkavara R kood.

1 Teststatistikud kuhja alternatiivhüpoteesi jaoks

Olgu populatsioon suurusega N jaotatud k -ks grupiks. Tähistame j -nda grupi vaatlused $x_{j1}, \dots, x_{jn_j}$, kus n_j on j -nda grupi vaatluste arv, seejuures $N = \sum_{j=1}^k n_j$.

Andmete kohta tehakse järgmised eeldused:

1. Vaatlused $x_{j1}, \dots, x_{jn_j}$ on sõltumatud iga $j = 1, \dots, k$ korral.
2. Iga fikseeritud $j = 1, \dots, k$ korral pärinevad vaatlused $x_{j1}, \dots, x_{jn_j}$ samast jaotusest jaotusfunktsiooniga F_j , kusjuures see jaotus on pidev.
3. Jaotusfunktsioonid $F_1, \dots, F_k$ erinevad üksteisest vaid nihke võrra, s.t iga $j = 1, \dots, k$ korral

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

kus F on mingi pideva jaotuse jaotusfunktsioon mediaaniga μ ja τ_j on j -nda grupi mõju uuritavale tunnusele.

4. Grupid $j = 1, \dots, k$ on järjestatud mõne tunnuse alusel, mille mõju uuritavale tunnusele soovitakse analüüsida.

(Hollander, Wolfe ja Chicken, [2014](#))

Selles peatükis esitatud testidega uuritakse hüpoteeside paari, kus nullhüpotees avaldub kujul

$$H_0 : F_1(x) \equiv \dots \equiv F_k(x)$$

ja alternatiivhüpotees avaldub kujul

$$H_1 : F_1(x) \geq \dots \geq F_{p-1}(x) \geq F_p(x) \leq F_{p+1}(x) \leq \dots \leq F_k(x),$$

kus vähemalt ühe x väärtuse korral on üks võrratustest range. Antud kujul esitatud alternatiivhüpoteesi nimetatatakse kuhja hüpoteesiks ning gruppi p kutsutakse

kuhja tipuks. (Mack ja Wolfe, 1981)

Järgnevalt vaadatakse lähemalt kolme erinevat testi, mis sobivad kuhja alternatiivhüpoteesi testimiseks. Kõik esitatud testid on esmalt disainitud olukorra jaoks, kus grupp p on enne hüpoteeside testimist teada, ning arendatud edasi olukorra jaoks, kus grupp p eelnevalt teada ei ole (Gökpınar ja Gökpınar, 2016).

1.1 Mack-Wolfe test

Mack-Wolfe test on kõige sagedamini kasutatav mitteparameetriline test kuhja alternatiivhüpoteeside jaoks (Gökpınar ja Gökpınar, 2016).

1.1.1 Teststatistik

Mack-Wolfe statistik on defineeritud Mann-Whitney statistiku kaudu. Olgu populatsioonis grupp u suurusega n_u ja grupp v suurusega n_v . Vaatleme kõiki paare (u_i, v_j) , kus u_i on i -s vaatlus grupis u ja v_j on j -s vaatlus grupis v . Mann-Whitney statistik gruppide u ja v vahel U_{uv} avaldub kujul

$$U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} I(u_i, v_j), \quad 1 \leq u < v \leq k, \quad (1.1)$$

kus

$$I(a, b) = \begin{cases} 1, & \text{kui } a < b \\ 0, & \text{kui } a \geq b \end{cases} \quad (1.2)$$

(Hollander, Wolfe ja Chicken, 2014)

Mack-Wolfe statistik on defineeritud järgnevalt:

$$MW_p = \sum_{1 \leq u < v \leq p} U_{uv} + \sum_{p \leq u < v \leq k} U_{vu}, \quad (1.3)$$

kus U_{uv} on u -nda ja v -nda grupi vaheline Mann-Whitney statistik 1.1. (Mack ja

Wolfe, 1981)

Sisuliselt hõlmab statistiku esimene liidetav endas kõiki Mann-Whitney statistikuid väiksema järjekorranumbri ja suurema järjekorranumbriga gruppide vahel ning seda kuni tipuni p . Teine statistiku liidetav sisaldab kõiki Mann-Whitney statistikuid peale kuhja tippu p , kusjuures sel puhul leitakse Mann-Whitney statistikud suurema järjekorranumbriga ja väiksema järjekorranumbriga gruppide vahel. Seega on oluline teadvustada, et üldiselt $U_{uv} \neq U_{vu}$.

Näide: teststatistiku väärtuse leidmine

Tahetakse uurida šokolaadi mõju tudengite vererõhu alanemisele. Olgu uuritavad jaotatud nelja gruppi. Igas grupis manustatakse osalejatele $50 \cdot x$ grammi šokolaadi, kus x on katses osaleja grupi number. Tabelis 1 on esitatud katses osalejate vererõhu langus tund aega peale šokolaadi söömist. Kokku osales uuringus 12 tudengit.

Tabel 1: Katsealuste vererõhu langus tund aega peale šokolaadi söömist.

Grupi number			
I	II	III	IV
4	11	12	8
9	14	18	13
7	6	15	10

Varasema uuringu kohaselt on optimaalne šokolaadi kogus vererõhu alandamiseks 150 grammi. Valime kuhja tipuks p grupi 3 ja leiame Mack-Wolfe statistiku.

Valemi 1.3 kohaselt avaldub Mack-Wolfe statistik Mann-Whitney statistikute summana

$$MW_3 = U_{12} + U_{13} + U_{23} + U_{43}.$$

Statistik U_{12} leitakse vastavalt valemitele 1.1 ja 1.2 järgnevalt:

$$\begin{aligned} U_{12} &= I(4, 11) + I(4, 14) + I(4, 6) + I(9, 11) + I(9, 14) + I(9, 6) + \\ &\quad + I(7, 11) + I(7, 14) + I(7, 6) = \\ &= 1 + 1 + 1 + 1 + 1 + 0 + 1 + 1 + 0 = 7 \end{aligned} \tag{1.4}$$

Sarnaselt leitakse ka ülejäänud Mann-Whitney statistikud

$$U_{13} = 9, \quad U_{23} = 8, \quad U_{43} = 8. \tag{1.5}$$

Seega on Mack-Wolfe statistik antud vaatluste ja kuhja tipu $p = 3$ valiku korral

$$MW_3 = 7 + 9 + 8 + 8 = 32.$$

Leitud Mack-Wolfe statistiku MW_3 põhjal saab langetada otsuse šokolaadi mõju kohta. Otsuse tegemiseks oleks vaja teada statistiku jaotust nullhüpoteesi kehtides ehk eeldusel, et kõik šokolaadikogused mõjuvad vererõhule sama moodi. Statistiku jaotuse leidmist käsitletakse järgmises alapeatükis.

1.1.2 Statistiku täpne jaotus nullhüpoteesi kehtides

Mitteparameetrilises statistikas kasutatakse tihti vaatluste endi asemel vaatluste astakuid. Näiteks kahe grupi võrdlemiseks kasutatav Mann-Whitney statistik loendab kokku, mitu korda esineb selline olukord, kus teise grupi vaatluse uuritava tunnuse väärtus on suurem kui esimese grupi vaatluse uuritava tunnuse väärtus. Sama tulemuse saab ka siis, kui kasutada vaatluste asemel vaatluste astakuid ühisest variatsioonreast, sest loendamisel on oluline vaid vaatluste omavaheline järjekord, kuid mitte vaatluste omavahelise erinevuse suurus.

Sellest tulenevalt ei muutu astakute kasutamisel ka Mack-Wolfe statistik 1.3, kuna

see on konstrueeritud Mann-Whitney statistikute kaudu. Seega Mack-Wolfe test-statistiku jaotuse leidmisel võib vaadata, millised on võimalikud astakute $1, \dots, N$ jagunemised k grupi vahel, kus gruppide mahud on vastavalt $n_1, \dots, n_k$.

Nullhüpoteesi kehtides on kõik astakute paigutused k grupis võrdvõimalikud. Kokku on selliseid astakute paigutusi

$$\frac{N!}{\prod_{j=1}^k n_j!}.$$

(Hollander, Wolfe ja Chicken, 2014)

Näitena on lisas A.1 toodud astakute kõikvõimalikud paigutused olukorras, kus $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ja $n_3 = 1$. Erinevaid võrdvõimalikke astakute paigutusi on sel juhul

$$\frac{5!}{2!2!1!} = 30.$$

Samas lisas on ära toodud ka Mack-Wolfe statistiku väärtused kõigi 30 erineva astakute paigutuse korral, mille põhjal saab leida statistiku täpse jaotuse. Saadud Mack-Wolfe statistiku täpne jaotus on esitatud tabelis 2.

Tabel 2: Mack-Wolfe statistiku MW_2 täpne jaotus nullhüpoteesi kehtides, kui $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ning $n_3 = 1$.

MW_2	0	1	2	3	4	5	6
p_i	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{3}{30}$	$\frac{3}{30}$

Statistiku täpse jaotuse abil saab leida, milline on tõenäosus näha nii ekstreemset või veel ekstreemsemat statistiku väärtust nullhüpoteesi kehtides, mis omakorda võimaldab testida peatüki 1 alguses esitatud hüpoteese.

Näide: Täpse jaotuse kasutamine hüpoteeside testimiseks

Jätkub näide šokolaadi söömise mõjust vererõhule. Tabeli 1 andmete põhjal tuli Mack-Wolfe statistiku väärtus 32. Statistiku erinevate väärtuste esinemiste sagedused nullhüpoteesi kehtides antud andmestiku korral on esitatud lisas A.2. Et leida tõenäosus saada konkreetne statistiku väärtus, tuleb vastava statistiku väärtuse esinemise sagedus läbi jagada kõikide astakute paigutuste koguarvuga. Olulisuse tõenäosus p arvutatakse järgnevalt:

$$\begin{aligned} p &= P(MW_3 \geq 32) = P(MW_3 = 32) + P(MW_3 = 33) + \\ &+ P(MW_3 = 34) + P(MW_3 = 35) + P(MW_3 = 36) = \\ &= \frac{1428}{369600} + \frac{840}{369600} + \frac{420}{369600} + \frac{168}{369600} + \frac{84}{369600} = 0,080. \end{aligned}$$

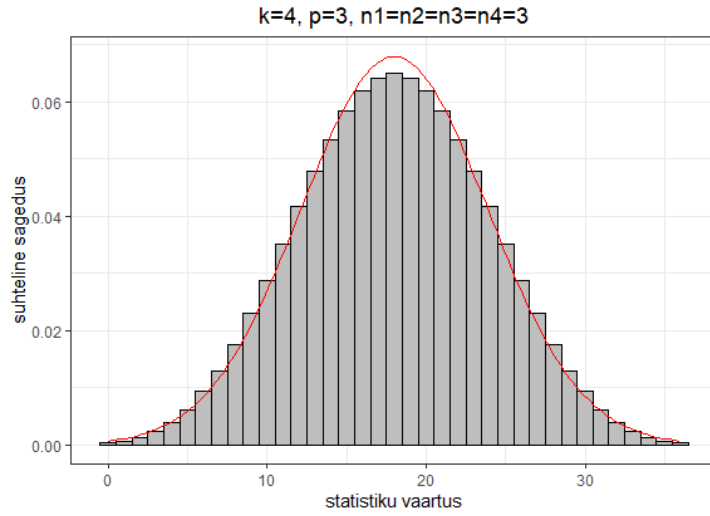
Kui olulisuse nivoo valida $\alpha = 0.05$, siis antud juhul võetakse vastu alternatiivhüpotees H_1 .

Samas on täpse jaotuse arvutamine vaatluste arvu suurenedes küllaltki tülikas protsess. Näiteks tabelis 1 toodud väikese näiteandmestiku korral on erinevaid astakute paigutusi kokku juba

$$\frac{12!}{3!3!3!} = 369600.$$

Seetõttu kasutatakse Mack-Wolfe statistiku põhjal otsuse tegemiseks enamasti statistiku asümptootilist jaotust. Joonisel 1 on näha, et Mack-Wolfe statistiku jaotus sarnaneb juba väikese andmestiku korral normaaljaotusega.

Tulbad joonisel tähistavad Mack-Wolfe statistiku täpset jaotust nullhüpoteesi kehtides ning punase joonega on tähistatud normaaljaotuse tihedusfunktsioon, mille keskväärtuse ja standardhälbe parameetrid on vastavalt kõikvõimalike astakute paigutuste põhjal leitud Mack-Wolfe statistikute väärtuste aritmeetiline keskmine ja standardhälve.



Joonis 1: Mack-Wolfe statistiku jaotus nullhüpoteesi kehtides.

1.1.3 Statistiku keskväärutus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides

Antud alapeatükis esitatud omaduste tõestused on leitavad artiklist (Archambault, Mack ja Wolfe, 1977).

Kui kehtib nullhüpotees ehk kui kõik vaatlused pärinevad samast jaotusest, siis Mack-Wolfe statistiku keskväärutus ja dispersioon avalduvad järgnevalt:

$$E_0(MW_p) = \left[N_1^2 + N_2^2 - \sum_{i=1}^k n_i^2 - n_p^2 \right] / 4 \quad (1.6)$$

$$D_0(MW_p) = \frac{1}{72} \left\{ 2(N_1^3 + N_2^3) + 3(N_1^2 + N_2^2) - \sum_{i=1}^k n_i^2(2n_i + 3) - n_p(2n_p + 3) + 12n_p N_1 N_2 - 12n_p^2 N \right\} \quad (1.7)$$

kus $N_1 = \sum_{j=1}^p n_j$ ja $N_2 = \sum_{j=p}^k n_j$. (Mack ja Wolfe, 1981)

Kui $\min(n_1, \dots, n_k) \longrightarrow \infty$, siis

$$MW_p^* = \frac{MW_p - E_0(MW_p)}{\sqrt{D_0(MW_p)}} \quad (1.8)$$

on asümptootiliselt standardse normaaljaotusega. (Gökpınar ja Gökpınar, 2016)

Näide: Asümptootilise jaotuse kasutamine hüpoteeside testimiseks

Asümptootilise jaotuse kasutamiseks on esmalt vaja leida teststatistiku keskväärtus ja dispersioon nullhüpoteesi kehtides. Šokolaadi söömise ja vererõhu seose hindamise näiteandmestiku (tabel 1) põhjal on need

$$\begin{aligned} E_0(MW_3) &= \left[9^2 + 6^2 - 4 \cdot 3^2 - 3^2 \right] / 4 = 18; \\ D_0(MW_3) &= \frac{1}{72} \left\{ 2(9^3 + 6^3) + 3(9^2 + 6^2) - \right. \\ &\quad \left. - 4 \cdot 3^2(2 \cdot 3 + 3) - 3 \cdot (2 \cdot 3 + 3) + 12 \cdot 3 \cdot 9 \cdot 6 - 12 \cdot 3^2 \cdot 12 \right\} = \\ &= \frac{2538}{72} = 35,25. \end{aligned}$$

Seejärel leitakse standardiseeritud statistiku väärtus

$$MW_3^* = \frac{MW_3 - E_0(MW_3)}{\sqrt{D_0(MW_3)}} = \frac{32 - 18}{\sqrt{35,25}} = 2,359$$

Testi olulisuse tõenäosuseks saadakse $(1 - \Phi(MW_3^*)) = 0,0092$, kus $\Phi(z)$ tähistab standardse normaaljaotuse jaotusfunktsiooni väärtust kohal z . Seega ka asümptootilise jaotuse kasutamisel võetakse antud näite korral vastu alternatiivhüpotees H_1 .

1.1.4 Kuhja tipu leidmine

Praktikas pole üldiselt eelnevat infot selle kohta, milline populatsiooni gruppidest on kuhja tipp p . Näiteks uue ravimi mõju uurimisel, kus ravimi saajad on gruppidesse jaotatud doosi suuruse alusel, ei tea uurijad ette, millise doosi korral on mõju patsientidele suurim. Sellisel juhul hinnatakse kõigepealt parim kandidaat kuhja tipu p jaoks ning seejärel leitakse Mack-Wolfe teststatistik MW_p alapeatükis 1.1.1 kirjeldatud meetodil. (Hollander, Wolfe ja Chicken, 2014)

Kuhja tipu leidmiseks arvutatakse esmalt valimi iga grupi $q = 1, \dots, k$ jaoks Mann-Whitney statistikute summa fikseeritud grupi ja kõikide ülejäänud valimi gruppide vahel

$$U_{.q} = \sum_{i \neq q} U_{iq}, \quad (1.9)$$

kus U_{iq} on Mann-Whitney statistik i -nda ja q -nda grupi vahel. Seejärel leitud väärtused standardiseeritakse ning saadakse statistik kujul

$$U_{.q}^* = \frac{U_{.q} - E_0(U_{.q})}{\sqrt{D_0(U_{.q})}}, \quad q = 1, \dots, k. \quad (1.10)$$

(Hollander, Wolfe ja Chicken, 2014)

Valemis 1.10 esitatud keskväärtus ja dispersioon avalduvad vastavalt

$$E_0(U_{.q}) = \frac{n_q(N - n_q)}{2} \quad (1.11)$$

$$D_0(U_{.q}) = \frac{n_q(N - n_q)(N + 1)}{12}, \quad (1.12)$$

kus n_q on grupi q vaatluste arv ja N on kogu valimi vaatluste arv. (Mack ja Wolfe, 1981)

Kuhja tipuks $\hat{p}$ valitakse grupp q , mille korral statistik $U_{.q}$ on suurim. Seejärel

arvutatakse Mack-Wolfe statistik alapeatükis 1.1.1 kirjeldatud meetodil, kasutades kuhja tipuna väärtust $\hat{p}$. (Hollander, Wolfe ja Chicken, 2014)

On võimalik, et suurim statistiku U_q väärtus on mitmel grupil korraga. Olgu kokku r gruppi, kus statistiku U_q on suurim ja olgu B alamhulk hulgast $\{1, \dots, k\}$. Hulka B kuuluvad nende gruppide indeksid, mis on suurima statistiku U_q väärtusega. Mack-Wolfe statistik arvutatakse sel juhul valemiga

$$MW_{\hat{p}}^* = \frac{1}{r} \sum_{j \in B} \left[\frac{MW_j - E_0(MW_j)}{\sqrt{D_0(MW_j)}} \right], \quad (1.13)$$

kus MW_j on Mack-Wolfe statistik, eeldusel, et kuhja tipp on grupp j . $E_0(MW_j)$ ja $D_0(MW_j)$ leitakse vastavalt peatükis 1.1.3 esitatud valemitele. (Hollander, Wolfe ja Chicken, 2014)

1.1.5 Võrdsed väärtused

Võib juhtuda, et vaadeldud väärtuste hulgas esineb võrdseid väärtusi. Sellisel juhul tuleks Mack-Wolfe statistiku 1.3 arvutamisel kasutada indikaatorfunktsiooni 1.2 asemel kuju

$$I(a, b) = \begin{cases} 1, & \text{kui } a < b \\ \frac{1}{2}, & \text{kui } a = b \\ 0, & \text{kui } a > b \end{cases} \quad (1.14)$$

Uuendatud kujul indikaatorfunktsiooni kasutamine ei mõjuta statistiku keskväärtust $E_0(MW_p)$ nullhüpoteesi kehtimisel. Küll aga väheneb statistiku dispersioon $D_0(MW_p)$ nullhüpoteesi kehtimisel, kui statistiku arvutamisel kasutada indikaatorfunktsiooni 1.14. (Hollander, Wolfe ja Chicken, 2014)

Kuna aga võrdsete väärtuste esinemisel pole täpse statistiku dispersiooni $D_0(MW_p)$ kohta nullhüpoteesi kehtimisel tõestatud tulemust, mis võimaldaks statistiku dispersiooni andmestiku parameetrite põhjal välja arvutada, siis kasutatakse disper-

siooni hinnanguna dispersiooni väärtust, mis on saadud, kui statistiku MW_p arvutamisel kasutada indikaatorfunktsiooni 1.2. Dispersiooni ülehindamise tõttu tuleb standardiseeritud statistik MW_j^* võrdsete vaatluste korral alati väiksem, kui täpse dispersiooni $D_0(MW_p)$ olemasolul. (Hollander, Wolfe ja Chicken, 2014)

Näide: Kuhja tipu leidmine võrdsete väärtuste korral

Selles alapeatükis vaadatakse näidet kuhja tipu määramisest, kui see eelnevalt teada pole. Lisaks esineb andmestikus võrdseid vaatlusi.

Kolmel identsel murulapil pandi kasvama samast sordist lilled. Muruplatsid nummerdati vastavalt 1, 2 ja 3. Igat muruplatsi kasteti iga päev $100 \cdot y$ milliliitri veega, kus y on muruplatsi number. Igal muruplatsil kasvatati kolme lille. Tabelis 3 on esitatud lillede varte pikkused sentimeetrites kuu aega pärast kasvama panemist.

Tabel 3: Lillede varte pikkused kuu aega peale kasvama panemist.

I	II	III
15	25	12
27	37	33
25	33	23

Kuhja tipu leidmiseks tuleb alapeatüki 1.1.4 põhjal arvutada $U_{.1}$, $U_{.2}$ ja $U_{.3}$.

$$U_{.1} = U_{21} + U_{31}$$

$$U_{.2} = U_{12} + U_{32}$$

$$U_{.3} = U_{13} + U_{23}$$

Kuna vaatluste hulgas esineb antud juhul võrdseid väärtusi, siis tuleb Mann-Whitney

statistikute arvutamisel kasutada indikaatorfunktsiooni [1.14](#).

$$\begin{aligned} U_{21} &= I(25, 15) + I(25, 27) + I(25, 25) + I(37, 15) + I(37, 27) + I(37, 25) + \\ &\quad + I(33, 15) + I(33, 27) + I(33, 25) = \\ &= 0 + 1 + \frac{1}{2} + 0 + 0 + 0 + 0 + 0 + 0 = \frac{3}{2} \end{aligned}$$

Samal põhimõttel leitakse ka teised Mann-Whitney statistikud:

$$U_{31} = 5, \quad U_{12} = \frac{15}{2}, \quad U_{32} = \frac{15}{2}, \quad U_{13} = 4, \quad U_{23} = \frac{3}{2}.$$

Seega statistikute

$$U_{.1} = 5 + \frac{3}{2} = \frac{13}{2}, \quad U_{.2} = \frac{15}{2} + \frac{15}{2} = \frac{30}{2}, \quad U_{.3} = 4 + \frac{3}{2} = \frac{11}{2}$$

põhjal on parim kandidaat kuhja tipu p jaoks grupp number 2.

1.2 Hettmansperger-Nortoni test

Hettmansperger ja Norton pakkusid välja oma testi kuhja alternatiivhüpoteesidele aastal 1987. Kogu alapeatükk [1.2](#) on kirjutatud Gökpinari ja Gökpinari 2016. aastal avaldatud artikli põhjal.

1.2.1 Teststatistik

Olgu R_{ji} j -nda grupi i -nda vaatluse x_{ji} astak kõigi N vaatluse hulgas. Olgu n_j j -nda grupi vaatluste arv ning olgu $\bar{R}_j = \frac{\sum_{i=1}^{n_j} R_{ji}}{n_j}$ j -nda grupi vaatluste astakute keskmine.

Kui kuhja tipp p on teada, siis olgu

$$c_j = \begin{cases} j, & \text{kui } j \leq p \\ 2p - j, & \text{kui } j > p. \end{cases} \quad (1.15)$$

Kordaja c_j näol on põhimõtteliselt tegemist kaaluga - kordaja on seda suurem, mida lähemal on grupp j kuhja tipule p .

Olgu ka

$$\lambda_j = \frac{n_j}{N} \quad \text{ja} \quad \bar{c}_w = \sum_{j=1}^k \lambda_j c_j.$$

Kordaja λ_j näitab grupi j vaatluste osakaalu kogu populatsioonis ning kordaja $\bar{c}_w$ on sisuliselt gruppide osakaaludega arvestav kordajate c_j keskmine.

Hettmansperger-Nortoni statistik avaldub kujul

$$HN_p = \sum_{j=1}^k \lambda_j (c_j - \bar{c}_w) \bar{R}_j. \quad (1.16)$$

Seega on antud statisiku HN_p arvutamisel arvestatud lisaks ka gruppide kaugustega kuhja tipust p , mida Mack-Wolfe statistiku MW_p korral ei tehtud.

Näide: Teststatistiku väärtuse leidmine

Näitena kasutatakse taas tabelis 1 toodud uuringu andmeid šokolaadi mõjust tungide vererõhule.

Iga $j \in \{1, 2, 3, 4\}$ korral on

$$n_j = 3, \quad \lambda_j = \frac{3}{12} = \frac{1}{4}.$$

Seega

$$\bar{c}_w = \sum_{j=1}^k \lambda_j c_j = \frac{1}{4} \sum_{j=1}^k (1 + 2 + 3 + 2) = 2.$$

Statistiku arvutamiseks tuleb leida ka iga grupi keskastakute väärtused, mis on esitatud tabelis 4.

Tabel 4: Katsealuste vererõhu muutuste astakud ja gruppide keskastakud.

Grupid	I	II	III	IV
Astakud	1	7	8	4
	5	10	12	9
	3	2	11	6
$\bar{R}_j$	3	$\frac{19}{3}$	$\frac{31}{3}$	$\frac{19}{3}$

Hettmannsperger-Nortoni statistik on antud vaatluste ja kuhja tipu $p = 3$ valiku korral valemi 1.16 põhjal

$$\begin{aligned}
 HN_3 &= \frac{1}{4} \cdot (1 - 2) \cdot 3 + \frac{1}{4} \cdot (2 - 2) \cdot \frac{19}{3} + \\
 &+ \frac{1}{4} \cdot (3 - 2) \cdot \frac{31}{3} + \frac{1}{4} \cdot (2 - 2) \cdot \frac{19}{3} = \\
 &= -\frac{3}{4} + 0 + \frac{31}{12} + 0 = \\
 &= \frac{22}{12}.
 \end{aligned}$$

1.2.2 Statistiku täpne jaotus nullhüpoteesi kehtides

Statistiku täpse jaotuse nullhüpoteesi kehtides saab ka Hettmannsperger-Nortoni statistiku jaoks leida jaotustabelina, vaadates läbi kõikvõimalikud astakute paigutused k gruppi.

Uurides näitena taas väga väikest andmestiku, kus $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ja $n_3 = 1$, on statistiku täpse jaotuse leidmine piisavalt kerge ka käsitsi arvutades. Kõik astakute paigutused ja vastavad statistiku väärtused on sarnaselt Mack-Wolfe statistikule esitatud lisas A.1. Tabelist 5 leiab statistiku täpse jaotuse antud näite korral.

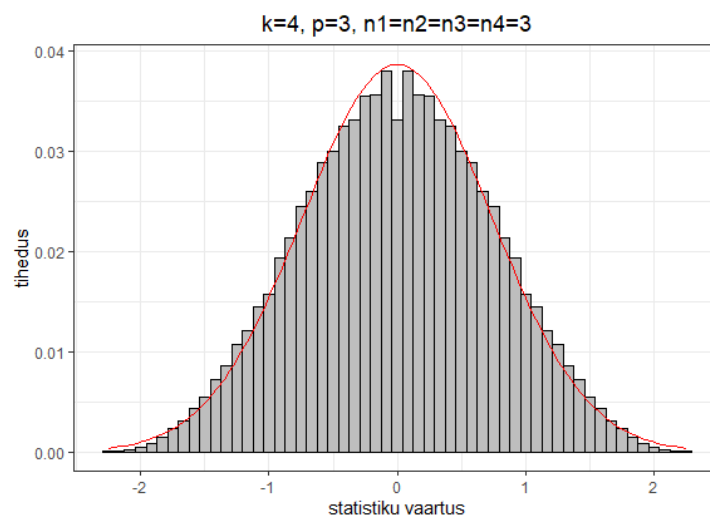
Antud andmestiku korral tulevad Hettmannsperger-Nortoni statistiku täpne jaotus

Tabel 5: Hettmannsperger-Nortoni statistiku täpne jaotus nullhüpoteesi kehtides, kui $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ja $n_3 = 1$.

HN_2	-0,6	-0,4	-0,2	0	0,2	0,4	0,6
p_i	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{3}{30}$	$\frac{3}{30}$

ning Mack-Wolfe statistiku täpne jaotus tabelis 2 kujult sarnased. Erinevalt Mack-Wolfe statistikust, mis omandab väärtustena vaid mittenegatiivseid täisarve, võib Hettmannsperger-Nortoni statistiku väärtus tulla ka negatiivne ning ei ole alati täisarv.

Kui võimalikke statistiku väärtusi on väga palju, siis saab statistiku täpset jaotust ülevaatlikult esitada histogrammi abil. Jooniselt 2 on näha, et Hettmannsperger-Nortoni statistik on eelnevalt toodud tabeli 1 näiteandmestiku korral sümmeetriline ja meenutab normaaljaotust. Seega pole üllatav, et on tõestatud ära tulemus statistiku asümptootilisest jaotusest.



Joonis 2: Hettmannsperger-Nortoni statistiku jaotus nullhüpoteesi kehtides.

Tulbad joonisel tähistavad Hettmannsperger-Nortoni statistiku täpset jaotust nullhüpoteesi kehtides ning punase joonega on tähistatud normaaljaotuse tihedusfunktsioon.

sioon, mille keskväärtuse ja standardhälbe parameetrid on vastavalt kõikvõimalike astakute paigutuste põhjal leitud Hettmannsperger-Nortoni statistikute väärtuste aritmeetiline keskmine ja standardhälve.

1.2.3 Statistiku keskväärtus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides

Antud alapeatükis esitatud omaduste tõestused on leitavad artiklist (Hettmannsperger ja Norton, 1987).

Kui kehtib nullhüpotees ehk kui kõik vaatlused pärinevad samast jaotusest, siis Hettmannsperger-Nortoni statistiku keskväärtus ja dispersioon avalduvad järgnevalt:

$$E_0(HN_p) = 0; \quad D_0(HN_p) = \frac{N+1}{12} \sum_{j=1}^k \lambda_j (c_j - \bar{c}_w)^2. \quad (1.17)$$

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

$$HN_p^* = \frac{HN_p}{\sqrt{D_0(HN_p)}} \quad (1.18)$$

on asümptootiliselt standardse normaaljaotusega.

Näide: Täpse ja asümptootilise jaotuse kasutamine hüpoteeside testimiseks

Tabeli 1 näiteandmestiku jaoks on Hettmannsperger-Nortoni statistiku täpne jaotus toodud lisas A.2. Teststatistiku arvutamise näite põhjal tuli antud andmestiku korral statistiku väärtus $\frac{22}{12}$.

Tõenäosus saada nii suur või veelgi suurem statistiku väärtus nullhüpoteesi kehtides täpse jaotuse põhjal on

$$\begin{aligned}
p &= P(HN_3 \geq \frac{22}{12}) = P(HN_3 = \frac{22}{12}) + P(HN_3 = \frac{23}{12}) + P(HN_3 = \frac{24}{12}) + \\
&+ P(HN_3 = \frac{25}{12}) + P(HN_3 = \frac{26}{12}) + P(HN_3 = \frac{27}{12}) = \\
&= \frac{600}{369600} + \frac{360}{369600} + \frac{200}{369600} + \frac{100}{369600} + \frac{40}{369600} + \frac{20}{369600} = 0,0036
\end{aligned}$$

Asümptootilise jaotuse kasutamiseks leitakse esmalt statistiku dispersioon antud andmestiku jaoks valemi 1.17 abil:

$$\begin{aligned}
D_0(HN_3) &= \frac{13}{12} \cdot \left[\frac{1}{4} \cdot (1-2)^2 + \frac{1}{4} \cdot (2-2)^2 + \frac{1}{4} \cdot (3-2)^2 + \frac{1}{4} \cdot (2-2)^2 \right] = \\
&= \frac{13}{24}
\end{aligned}$$

Seejärel leitakse standardiseeritud statistik valemi 1.18 põhjal:

$$HN_3^* = \frac{HN_3}{\sqrt{D_0(HN_3)}} = \frac{\frac{22}{12}}{\sqrt{\frac{13}{24}}} = 2,491.$$

Testi olulisuse tõenäosuseks saadakse $1 - \Phi(HN_3^*) = 0,0064$, kus $\Phi(z)$ tähistab standardse normaalkaotuse jaotusfunktsiooni väärtust kohal z .

Seega antud juhul võetakse nii täpse kui ka asümptootilise jaotuse kasutamisel olisuse nivoo $\alpha = 0,05$ korral vastu alternatiivhüpotees H_1 .

Täpse jaotusega ning asümptootilise jaotusega leitud p -väärtused tulid küll selles näites üsna erinevad, kuid tuleb ka silmas pidada, et valimimaht oli siin näite puhul veel üsna väike ($N = 12$).

1.2.4 Kuhja tipu leidmine

Kui kuhja tipp p pole teada, siis omandab Hettmansperger-Nortoni standardiseeritud statistik kuju

$$HN_{\max}^* = \max(HN_1^*, \dots, HN_k^*), \quad (1.19)$$

kus HN_j^* , $j = 1, \dots, k$ on arvutatud valemi 1.18 järgi. Teisisõnu valitakse kuhja tipuks $\hat{p}$ grupp j , mille korral tuleb standardiseeritud statistiku väärtus suurim.

Hüpoteeside testimiseks oleks vaja teada statistiku $HN_{\max}^*$ jaotust. Kuna see pole lihtsasti leitav, kasutatakse p -väärtuse leidmiseks järgnevat asümptootilist tulemust.

Saab näidata, et nullhüpoteesi kehtides on k -mõõtmelise vektori

$$H\vec{N}^* = (HN_1^*, \dots, HN_k^*)^T,$$

mille elemendid on standardiseeritud Hettmannsperger-Nortoni statistikud HN_j^* , asümptootiliseks jaotuseks mitmemõõtmeline normaaljaotus keskväärtusvektoriga $(0, \dots, 0)^T$ ja kovariatsioonimaatriksiga Σ . Maatriksi Σ elemendid avalduvad kujul:

$$\sigma_{st} = Cov_0(HN_s^*, HN_t^*) = \frac{\sum_{j=1}^k \lambda_j (c_{sj} - \bar{c}_{sw})(c_{tj} - \bar{c}_{tw})}{\sqrt{\sum_{j=1}^k \lambda_j (c_{sj} - \bar{c}_{sw})^2 \sum_{j=1}^k \lambda_j (c_{tj} - \bar{c}_{tw})^2}}.$$

Genereerides juhuslikke vektoreid selles k -mõõtmelisest normaaljaotusest saab hinnata ka vektori elementide maksimumi jaotust ja leida sel viisil testi p -väärtust.

1.3 Modifitseeritud Mack-Wolfe statistik

E.Gökpınar ja F.Gökpınar esitasid 2016. aastal modifitseeritud kuju Mack-Wolfe statistikule, mis arvestab statistiku leidmisel ka kahe grupi omavahelise kaugusega. Alapeatükk 1.3 on kirjutatud Gökpınari ja Gökpınari 2016. aastal avaldatud artikli põhjal.

1.3.1 Teststatistik

Modifitseeritud Mack-Wolfe statistik avaldub kujul

$$MMW_p = \sum_{i=1}^{p-1} \sum_{j=i+1}^p (j-i)U_{uv} + \sum_{i=p}^{k-1} \sum_{j=i+1}^k (j-i)U_{uv} \quad (1.20)$$

kus U_{uv} on u -nda ja v -nda grupi vaheline Mann-Whitney statistik.

Sisuliselt erineb modifitseeritud statistik originaalsest selle poolest, et kahe grupi vahelised Mann-Whitney statistikud on läbi korrutatud gruppide omavahelise kaugusega.

Näide: Teststatistiku väärtuse leidmine

Näitena kasutatakse taas tabelis 1 toodud vaatlusi ning statistiku arvutamisel lähutatakse teadmisest, et kuhja tipp $p = 3$.

Modifitseeritud Mack-Wolfe statistik arvutatakse valemi 1.20 põhjal

$$MMW_3 = (2-1) \cdot U_{12} + (3-1) \cdot U_{13} + (3-2) \cdot U_{23} + (4-3) \cdot U_{43}.$$

Varasemast 1.4 ja 1.5 põhjal on juba teada, et

$$U_{12} = 7, \quad U_{13} = 9, \quad U_{23} = 8, \quad U_{43} = 8.$$

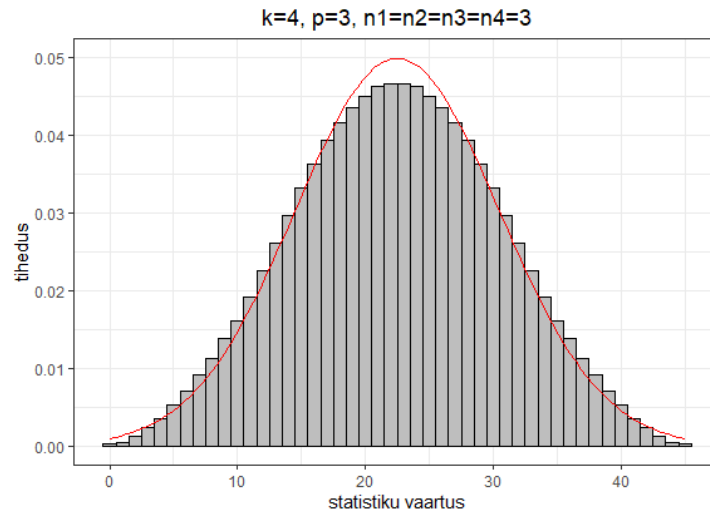
Seega Modifitseeritud Mack-Wolfe statistik on valemi 1.20 põhjal

$$MMW_3 = 7 + 2 \cdot 9 + 8 + 8 = 41.$$

1.3.2 Statistiku täpne jaotus nullhüpoteesi kehtides

Teststatistiku täpne jaotus on väga väikese näiteandandmestiku korral, kus $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ning $n_3 = 1$, toodud lisas A.1.

Nagu ka Mack-Wolfe statistiku puhul, tulevad modifitseeritud Mack-Wolfe statistiku väärtused alati täisarvud. Tabelis 1 toodud andmete korral on statistiku väärtuste esinemiste sagedused esitatud lisas A.2. Joonisel 3 on esitatud histogramm antud näite statistiku väärtuste jagunemisest.



Joonis 3: Modifitseeritud Mack-Wolfe statistiku jaotus nullhüpoteesi kehtides.

Tulbad joonisel tähistavad modifitseeritud Mack-Wolfe statistiku täpset jaotust nullhüpoteesi kehtides ning punase joonega on tähistatud normaaljaotuse tihedusfunktsioon, mille keskväärtuse ja standardhälbe parameetrid on vastavalt kõikvõimalike astakute paigutuste põhjal leitud modifitseeritud Mack-Wolfe statistikute väärtuste aritmeetiline keskmine ja standardhälve.

Sarnaselt eelnevale kahele statistikule on ka modifitseeritud Mack-Wolfe statistik nullhüpoteesi kehtides sümmeetrilise jaotusega, mis meenutab normaaljaotust. Ka antud statistiku jaoks leidub tulemus statistiku asümptootilisest jaotusest, mis lihtsustab hüpoteeside uurimist suuremate valimimahtude korral.

1.3.3 Statistiku keskvärtus, dispersioon ja asümptootiline jaotus nullhüpoteesi kehtides

Antud alapeatükis esitatud omaduste tõestused on leitavad artiklist (Gökpınar ja Gökpınar, 2016).

Kui gruppide mahud $n_1, \dots, n_k$ on võrdse suurusega n , siis modifitseeritud Mack-Wolfe statistiku keskvärtus ja dispersioon avalduvad nullhüpoteesi kehtides järgnevalt:

$$E_0(MMW_p) = \frac{n^2}{2} \left[\binom{p+1}{3} + \binom{k-p+2}{3} \right] \quad (1.21)$$

$$\begin{aligned} D_0(MMW_p) = & \\ &= \frac{n^2(k-p+1)^2((k-p+1)^2-1)(n(k-p+1)+1)}{144} + \\ &+ \frac{n^2p^2(p^2-1)(np+1)}{144} + \frac{n^3p(p-1)(k-p)(k-p+1)}{24} \end{aligned} \quad (1.22)$$

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

$$MMW_p^* = \frac{MMW_p - E_0(MMW_p)}{\sqrt{D_0(MMW_p)}} \quad (1.23)$$

on asümptootiliselt standardse normaaljaotusega.

Näide: Täpse ja asümptootilise jaotuse kasutamine hüpoteeside testimiseks

Tabelis 1 toodud šokolaadi söömise mõju hindamise näite andmete põhjal arvutatud modifitseeritud Mack-Wolfe statistiku täpne jaotus on esitatud lisas A.2. Modifitseeritud Mack-Wolfe statistiku väärtus tuli näiteandmestiku korral 41.

Tõenäosus saada nullhüpoteesi kehtides nii suur või veelgi suurem statistiku väärtus täpset jaotust kasutades on

$$p = P(MMW_3 \geq 41) = P(MMW_3 = 41) + P(MW_3 = 42) + \\ + P(MW_3 = 43) + P(MW_3 = 44) + P(MW_3 = 45) = 0,0076$$

Asümptootilise jaotuse kasutamiseks leitakse esmalt statistiku keskväärts ja dispersioon antud andmestiku ($k = 4$, $p = 3$, $n = n_1 = n_2 = n_3 = n_4 = 3$) jaoks valemite 1.21 ja 1.22 põhjal:

$$E_0(MMW_3) = \frac{3^2}{2} \cdot \left[\binom{4}{3} + \binom{3}{3} \right] = \frac{9}{2} \cdot 5 = \frac{45}{2}; \\ D_0(MMW_3) = \frac{3^2 \cdot 2^2 \cdot (2^2 - 1) \cdot (3 \cdot 2 + 1)}{144} + \\ + \frac{3^2 \cdot 3^2 \cdot (3^2 - 1) \cdot (9 + 1)}{144} + \frac{3^3 \cdot 3 \cdot 2 \cdot 1 \cdot 2}{24} = \\ = \frac{7326}{144} + \frac{324}{24} = 64,375.$$

Seejärel leitakse standardiseeritud statistik valemi 1.23 põhjal:

$$MMW_3^* = \frac{MMW_3 - E_0(MMW_3)}{\sqrt{D_0(MMW_3)}} = \frac{41 - \frac{45}{2}}{\sqrt{64,375}} = 2,306.$$

Testi olulisuse tõenäosuseks saadakse $1 - \Phi(MMW_3^*) = 0,011$, kus $\Phi(z)$ tähistab standardse normaalfaotuse jaotusfunktsiooni väärtust kohal z .

Seega antud juhul võetakse nii täpse kui ka asümptootilise jaotuse kasutamisel olisuse nivoo $\alpha = 0.05$ korral vastu alternatiivhüpotees H_1 .

1.3.4 Kuhja tipu leidmine

Kui kuhja tipp p pole teada, siis rakendatakse sama ideed, mis tavalise Mack-Wolfe testi puhul alapeatükis 1.1.4. Kuhja tipp $\hat{p}$ hinnatakse statistiku $U_{\hat{p}} = \max(U_{.1}^*, \dots, U_{.k}^*)$ põhjal, kus $U_{.j}^*$, $j = 1, \dots, k$ on leitud valemiga 1.10.

Kuhja tipuks $\hat{p}$ valitakse grupp q , mille korral statistik $U_{.q}$ on suurim. Modifitseeritud Mack-Wolfe statistik arvutatakse seejärel valemiga [1.20](#), kasutades kuhja tipuna väärtust $\hat{p}$. Testi tegemiseks leitakse statistiku täpne jaotus alapeatükis [1.3.2](#) kirjeldatud viisil. Teine võimalus on statistik standardiseerida valemi [1.23](#) kohaselt ning kasutada statistiku asümptootilist jaotust nagu on kirjeldatud peatükis [1.3.3](#).

2 Testide võrdlemine

Selles peatükis uuritakse eelnevalt kirjeldatud Mack-Wolfe, Hettmannsperger- Nortoni ning modifitseeritud Mack-Wolfe testide käitumist simulatsioonide abil. Võrdluseks on juurde lisatud ka Kruskal-Wallise test, mis on üldine mitteparameetriline test k grupi erinevuse kontrollimiseks. Genereeritud andmete põhjal hinnatakse I liiki vea tegemise tõenäosust ja testi võimsust erinevate valimimahtude korral, kui kasutada otsuse tegemiseks testi asümptootilist jaotust. Simulatsioonid on läbi viidud erinevate gruppide arvu k ja tipu p paiknemise korral. Vaadatud on nii sümmeetrilisest kui ka ebasümmeetrilisest jaotusest pärinevaid andmeid.

Iga valitud parameetrite komplektiga korratakse andmestiku genereerimist ja testide läbiviimist 10000 korda. Testi otsuse tegemiseks kasutatakse olulisuse nivood $\alpha = 0,05$.

Andmete genereerimine ning testide tegemiseks vajalikud arvutused viiakse läbi statistikatarkvaras R (R Core Team, 2021). Tulemused on esitatud graafikutena alapeatükkides 2.2 ja 2.3.

Funktsioonid andmete genereerimiseks on esitatud lisas A.3, funktsioonid statistikute väärtuste leidmiseks lisas A.4, funktsioonid testide tegemiseks lisas A.5 ning funktsioonid jooniste trükkimiseks lisas A.6.

2.1 Andmete genereerimine

Võrdlemiseks kasutatavate andmete genereerimisel tuleb silmas pidada peatüki 1 alguses välja toodud eeldusi. Töös kasutatud funktsioonid andmete tekitamiseks on loodud nii, et iga grupi vaatlused on genereeritud vastavalt R programmi käsuga `rnorm` või `rexp`, mis genereerivad vastavalt etteantud parameetritega normaaljaotusega või eksponentjaotusega juhuslikke suursusi. See tagab eelduste 1 ja 2 kehtimise, ehk et iga fikseeritud grupi vaatlused on sõltumatud ning pärinevad samast pidevast jaotusest.

Eelduse 4 võib lugeda kehtivaks ülesande püstitusega, st. kuna kasutatakse simuleeritud vaatlusi, siis eeldame, et need vaatlused pärinevad uuringust, kus grupid olid järjestatud mingi tunnuse alusel, mille mõju uuritavale tunnusele tahetakse uurida.

Eelduse 3 täitmiseks peavad jaotuste jaotusfunktsioonid, kust andmed on gruppidesse genereeritud, erinema üksteisest vaid nihke võrra. Antud töös vaadatakse kahte juhtu, kus see eeldus on täidetud, kuid lisaks uuritakse ka olukorda, kus toimub lisaks nihkele toimub muutus ka hajuvuses, et hinnata testide käitumist olukorras, kus see eeldus pole täidetud. Seega on andmete genereerimiseks kasutatud kolme erinevat meetodit:

1. Igas grupis genereeritakse normaaljaotusega andmed, kus standardhälve parameeter σ on fikseeritud ning gruppide vahel varieeritakse normaaljaotuste keskväärtuste parameetreid.
2. Igas grupis genereeritakse eksponentjaotusega andmed, kus parameeter λ on fikseeritud ning gruppides on andmeid nihutatud valitud konstantide võrra.
3. Igas grupis genereeritakse eksponentjaotusega andmed, kus gruppide vahel varieeritakse parameetri λ väärtusi.

Variantide 1 ja 2 puhul on eeldus 3 täidetud. Variant 3 korral see aga nii ei ole, sest eksponentjaotuse parameetrit λ varieerides muutuvad ka jaotuste hajuvused. See variant on võetud kasutusse, et uurida, kuidas käituvad testid eelduse 3 mittekehtimisel.

Lisaks andmete genereerimise meetodi valimisele on testandmestiku tekitamiseks vaja fikseerida ka gruppide arv k , kuhja tipu asukoht p ning vaatluste arv igas grupis n_j , $j = 1, \dots, k$. Antud töös on vaatluste arv igas grupis võetud sama ehk $n_j = n$, $j = 1, \dots, k$, sest standardiseeritud modifitseeritud Mack-Wolfe statistiku asümptootilise jaotuse tulemuse 1.23 kehtimiseks on see eeldus vajalik.

Parameetreid k , p ja n varieeritakse käesolevas töös järgnevalt:

- Gruppide arv $k \in \{3, 4, 5, 6, 7\}$
- Kuhja tipp $p \in \{1, \dots, k\}$
- Vaatluste arv grupis $n \in \{5, 10, 15, \dots, 45, 50\}$

Võimsuse hindamisel on iga gruppide arvu k korral läbi vaadatud kõikvõimalikud kuhja tipu p paiknemise juhud. Iga k ja p kombinatsiooni korral on genereeritud andmed kolme eeltoodud jaotuse järgi (st. olukorras, kus kehtib alternatiivne hüpotees) ja leitud iga testi korral sisuka hüpoteesi vastuvõtmise suhteline sagedus.

Esimest liiki vea tegemise tõenäosuse hindamiseks on andmed gruppides genereeritud kas samast normaaljaotusest või samast eksponentjaotusest, et võrrelda, kas I liiki vea tegemise tõenäosustes on erinevus sümmeetrilise ja ebasümmeetrilise jaotuse vahel. Läbi on vaadatud juhud, kus $k = 4$, $p \in \{1, 2\}$ ning $k = 7$, $p \in \{1, 2, 3\}$.

2.2 Võimsuste võrdlemine

Järgnevalt analüüsitakse testide võimsusi graafikute abil. Kõikidel joonistel on x -teljel välja toodud gruppide $n_1, \dots, n_k$ maht n ning y -teljel vastavalt kui suures osas test 10000 genereeritud andmestikust võttis vastu alternatiivse hüpoteesi H_1 olulisuse nivool $\alpha = 0,05$. Neli uuritavat testi on esitatud eri värvi ja joonetüübiga.

Teststatistikute sümmeetrilisuse tõttu on iga fikseeritud k korral juhud $p = i$ ja $p = k - i + 1$, kus $i \in \{1, \dots, \lceil \frac{k}{2} \rceil\}$, võrdväärsed. Siinkohal on välja toodud vaid joonised neist esimese juhu jaoks.

Seega $k = 3$ ja $k = 4$ kohta on esitatud joonised, kus $p \in \{1, 2\}$, $k = 5$ ja $k = 6$ kohta joonised, kus $p \in \{1, 2, 3\}$, ning $k = 7$ kohta joonised, kus $p \in \{1, 2, 3, 4\}$.

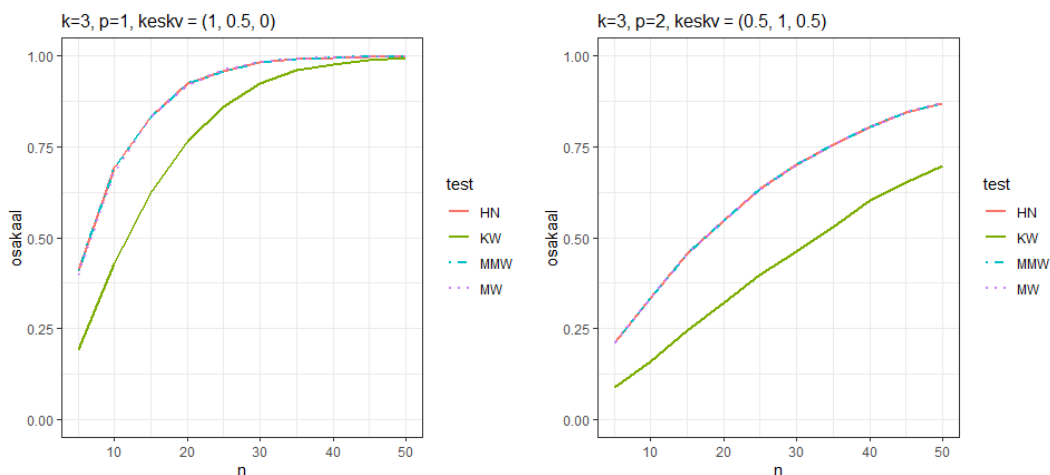
Lihtsuse huvides lühendatakse testide nimed Mack-Wolfe, Hettmannsperger-Nortoni, modifitseeritud Mack-Wolfe ja Kruskal-Wallise vastavalt MW, HN, MMW ja KW.

2.2.1 Normaaljaotusega andmed

Selles alapeatükis pärinevad fikseeritud k korral iga grupi $1, \dots, k$ jaoks vaatlused erineva keskväärtuse parameetriga μ normaaljaotustest. Järgitud on põhimõtet, et kuhja tipuks valitud grupi p korral oleks μ suurim. Grupile p eelnevate ja järgnevate gruppide normaaljaotuse keskväärtuse parameetrit μ on vähendatud konstandi võrra. Keskväärtuste parameetrid kõikides gruppides on välja toodud vastavate jooniste pealkirjade juures. Standardhälve parameeter σ on igal pool võetud 1.

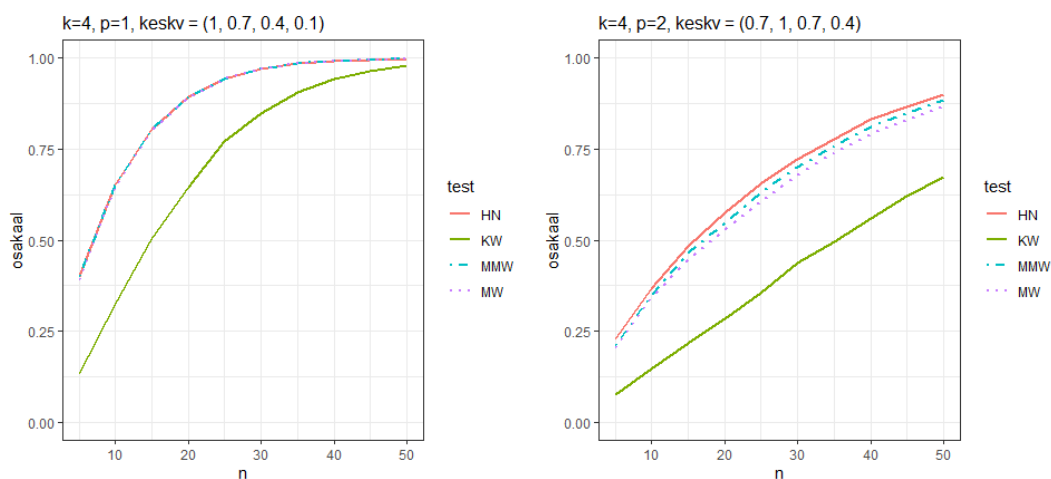
Joonisel 4 on näha, et kui $k = 3$, siis nii $p = 1$ kui ka $p = 2$ korral käituvad kõik kolm kuhja alternatiivhüpoteeside jaoks kirjeldatud testi MW, HN ja MMW identselt. MW ja MMW testide puhul on $p = 2$ korral see tulemus oodatav, sest sellisel juhul on MMW statistik on definitsioonist 1.20 tulenevalt võrdne MW statistikuga.

Ootuspäraselt on KW test väiksema võimsusega, kuna see ei kasuta eelnevat infot andmete kuhjakujulisusest gruppide vahel. Kusjuures KW test töötab paremini kuhja erijuhul ehk siis, kui kuhja tipp asub kõige äärmises grupis, sest sel juhul on tegu monotoonselt kahanevalt järjestatud gruppidega.



Joonis 4: Testide võimsused normaaljaotusega andmete korral, $k = 3$

Joonise 5 põhjal saab väita, et kui $k = 4$, siis $p = 1$ korral on MW, HN ja MMW testid endiselt sama käitumisega, edestades KW testi iga uuritud n juures. Kui $p = 2$, siis on märgata väikest erinevust testide võimsuste vahel, kus iga uuritud n juures on suurima võimsusega HN test, millele järgnevad MMW ning MW testid. Sarnaselt $k = 3$ juhuga on siingi kõikide testide võimsused iga n väärtuse korral suuremad kuhja tipu erijuhul, kui kuhja tipp asub kõige äärmises grupis, ehk siis, kui $p = 1$.

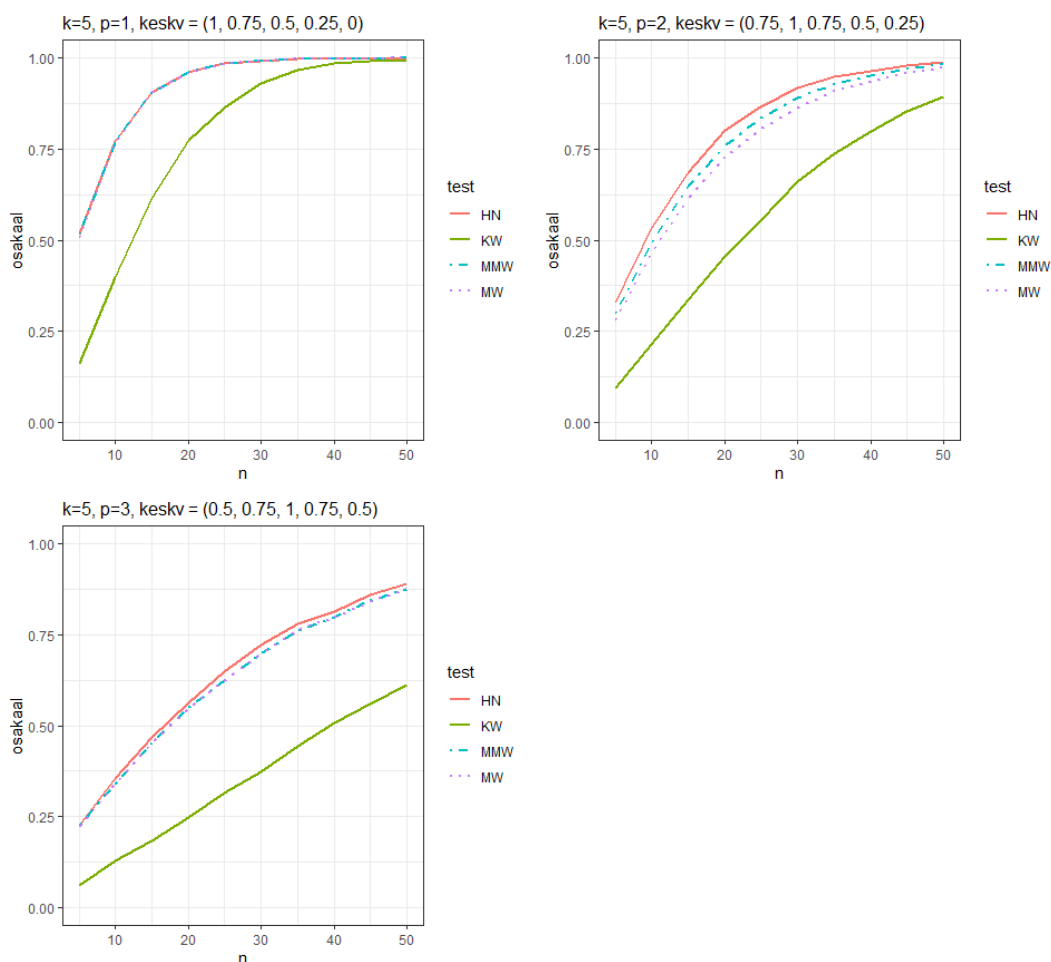


Joonis 5: Testide võimsused normaaljaotusega andmete korral, $k = 4$

Sarnast juhtu on vaadatud ka Gökpinari ja Gökpinari 2016. aastal avaldatud artiklis. Seal esitatud analüüsis jõuti teistsugusele tulemusele - HN testi võimsus tuli $k = 4$ ja $p = 2$ korral madalam, kui MW ja MMW testide võimsused, mis olid kõikide n väärtuste korral omavahel võrdsed. See võib olla tingitud asjaolust, et artiklis kasutati teistsuguseid normaaljaotuste keskväärtuste parameetreid gruppide vaatluste genereerimisel, mis olid vastavalt $(0,25; 1; 0,25; 0,1)$.

Jooniselt 6 nähtub, et kui $k = 5$, siis $p = 1$ korral jätkub sama muster, kus MW, HN ja MMW testid käituvad identselt ning KW test suudab väiksema valimimahu n korral märgatavalt vähematel juhtudel võtta vastu alternatiivse hüpoteesi H_1 . Kui $p = 2$, siis HN test on jällegi iga n juures suurima võimsusega, edestades vähesel

määral MMW ning MW teste. Sama kehtib ka $p = 3$ korral, kuid siin ei eristu enam MW ja MMW testid. KW test on nii $p = 2$ kui ka $p = 3$ juhtudel tunduvalt väiksema võimsusega kui teised testid.

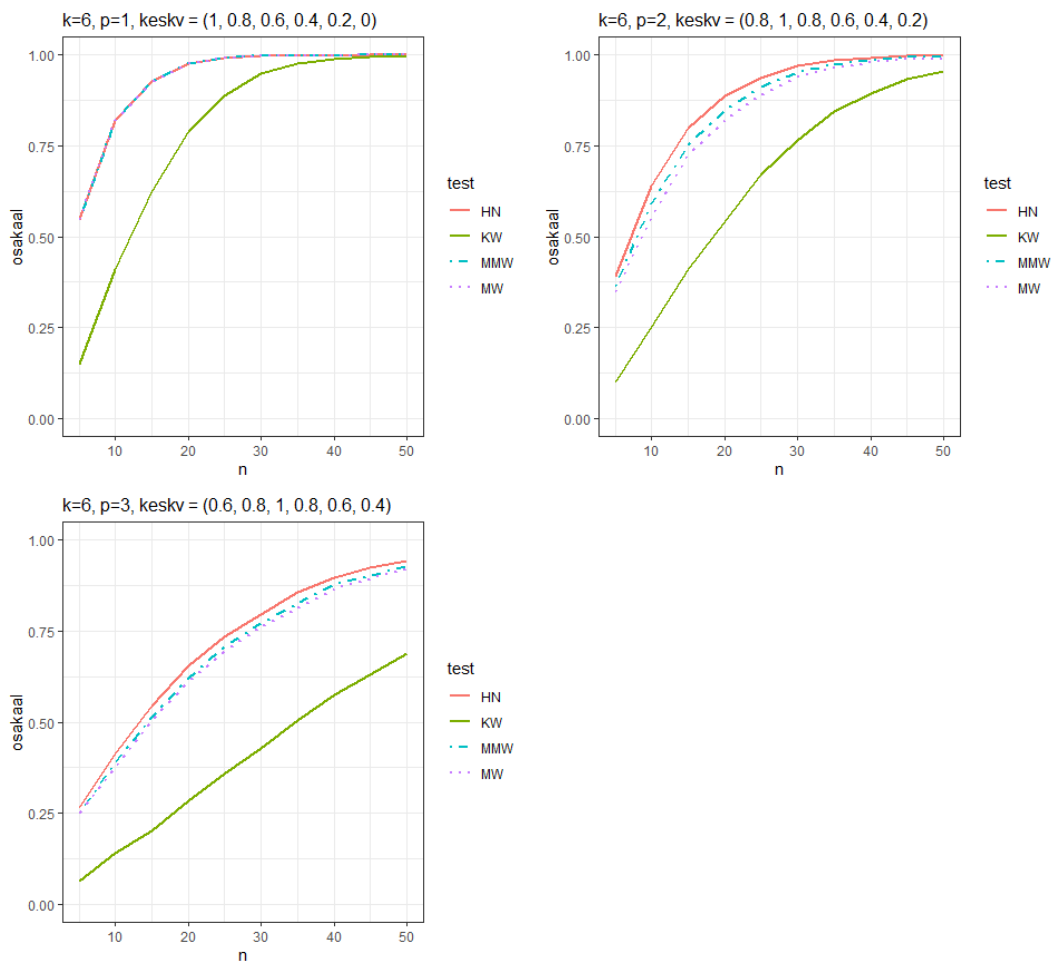


Joonis 6: Testide võimsused normaaljaotusega andmete korral, $k = 5$

Variandi korral, kus $k = 5$ ja $p = 3$, said Gökpınarid samuti tulemuseks, et HN test tuli võimsuselt väiksem kui MW ja MMW testid. Taaskord on see ilmselt tingitud asjaolust, et kasutatavad normaaljaotuste keskväärtuste parameetrid, kust gruppide vaatlusi genereeriti, olid kuhja tipu eristumise suhtes järsemad, vastavalt $(0,1; 0,25; 1; 0,25; 0,1)$. Juhul, kui $p = 2$, tulid selles artiklis MW, HN ja MMW testide võimsused võrdsed, mis samuti ei lange kokku käesolevas töös saadud tu-

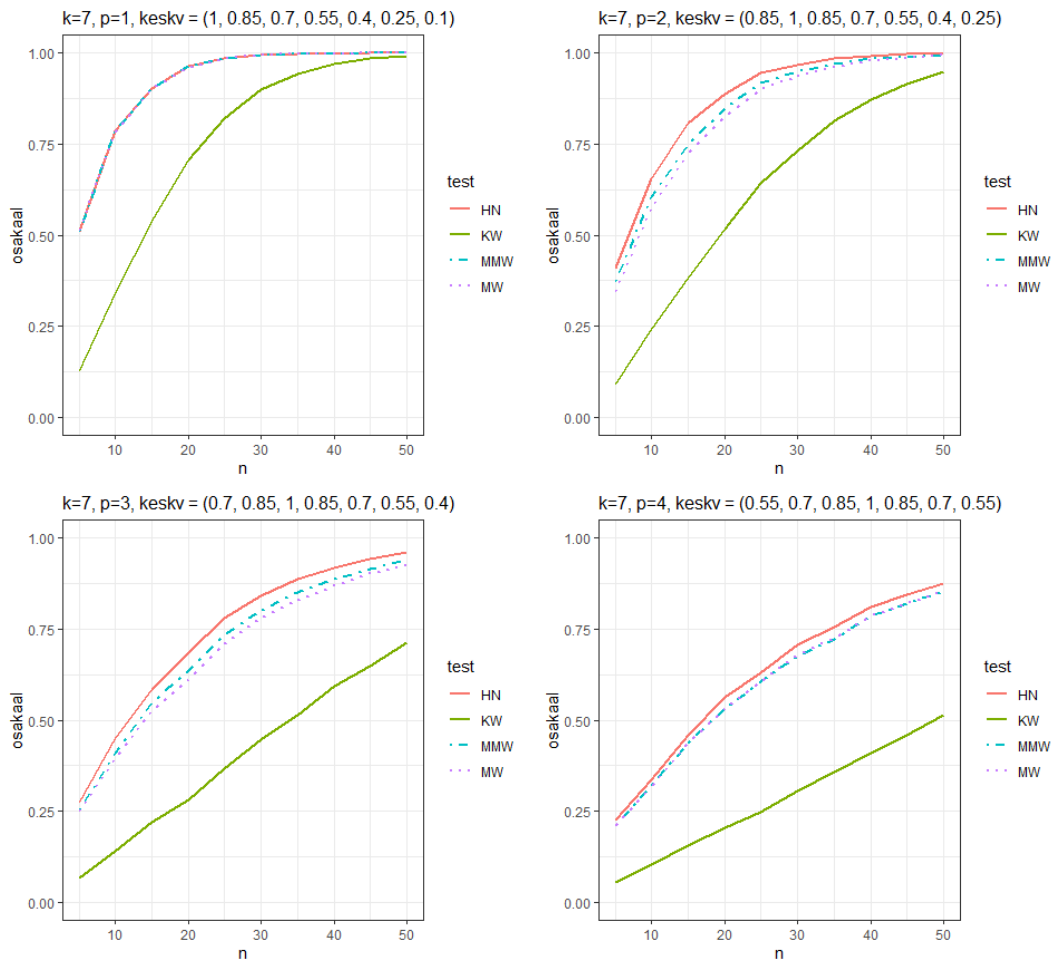
lemustega. Seega on alust arvata, et nimetatud testide võimsused sõltuvad lisaks lähtejaotusele ka nihete suurustest. Tasub ka märkida, et Gökpinarite artiklis pole öeldud, millised olid gruppide vaatluste genereerimiseks kasutatavate normaaljaotuste standardhälvete parameetrid, mis võib samuti saadud tulemust mõjutada.

Joonise 7 põhjal saab öelda, et kui $k = 6$, siis erinevate p väärtuste korral on testide käitumine sarnane $k = 5$ juhule. Ainsa erinevusena eristuvad seekord $p = 3$ juures MW ja MMW testid, kuid vahe näib olevat minimaalne. Tuleb silmas pidades, et erinevalt $k = 5$ juhust ei vasta siinkohal $p = 3$ sümmeetrilisele kuhjale.



Joonis 7: Testide võimsused normaaljaotusega andmete korral, $k = 6$

Jooniselt 8 on näha, et kui $k = 7$, siis $p = 1$ korral MW, HN ja MMW testid endiselt omavahel ei eristu. Suurimad erinevused võimsuste juures on $p = 2$ ning $p = 3$ korral, kus HN test tuvastab kõige rohkematel juhtudel alternatiivse hüpoteesi H_1 kehtimise. Võimsuselt teisele kohale paigutub MMW test, mis edestab väga vähesel määral MW testi. KW testi võimsus jääb teistele kolmele testile märgatavalt alla, kusjuures näib, et vahe teiste testidega on seda suurem, mida keskemal on kuhja tipp p .

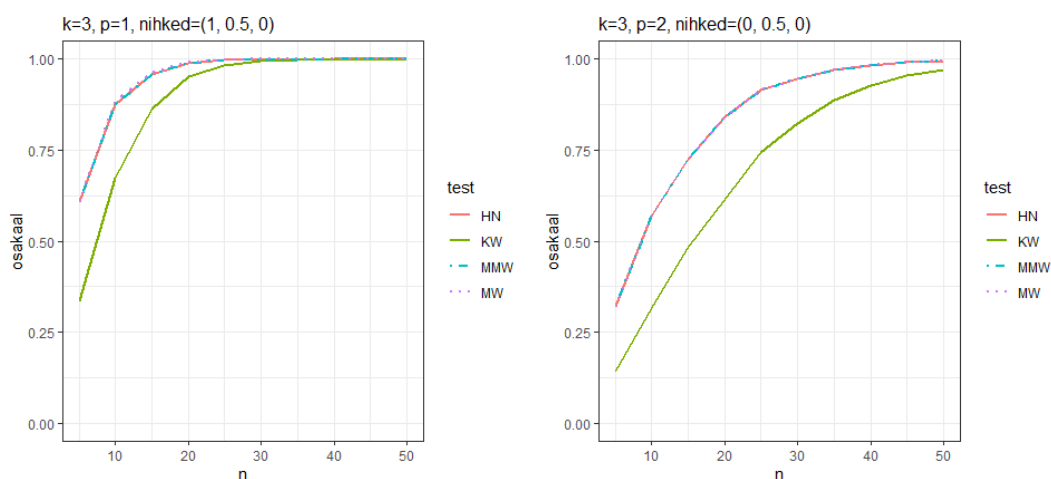


Joonis 8: Testide võimsused normaaljaotusega andmete korral, $k = 7$

2.2.2 EkspONENTJAOTUSEGA ANDMED NIHUTATUD KONSTANDI VÕRRA

Antud alapeatükis pärinevad fikseeritud k korral gruppide $1, \dots, k$ vaatlused parameetriga $\lambda = 1$ eksponentjaotusest. Peale andmete genereerimist on iga grupi vaatlusi nihutatud teatud konstandi võrra. Järgitud on põhimõtet, et vaatlused paikneksid gruppide vahel kuhja kujuga. See tähendab, et suurima konstandiga on nihutatud kuhja tipuks valitud grupi p vaatlusi. Grupile p eelnevatele ja järgnevaatele gruppidele on nihke teostamiseks kasutatud konstanti vähendatud fikseeritud sammuga.

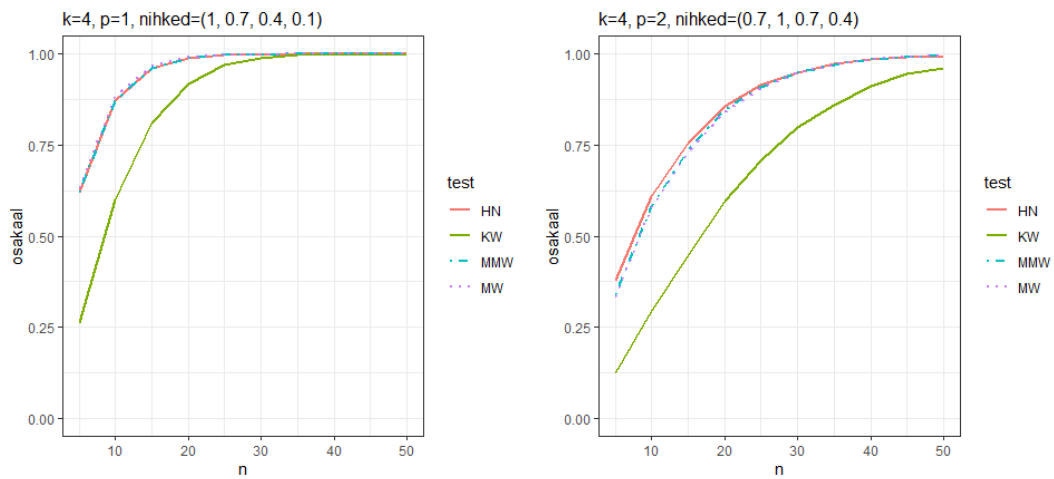
Joonise 9 põhjal võib väita, et kui $k = 3$, siis $p = 1$ ja $p = 2$ korral käituvad MW, HN ning MMW testid identseks. KW test on ootuspäraselt ülejäänud testidest madalama võimsusega.



Joonis 9: Testide võimsused eksponentjaotusega andmete korral, mida on nihutatud konstandi võrra, $k = 3$

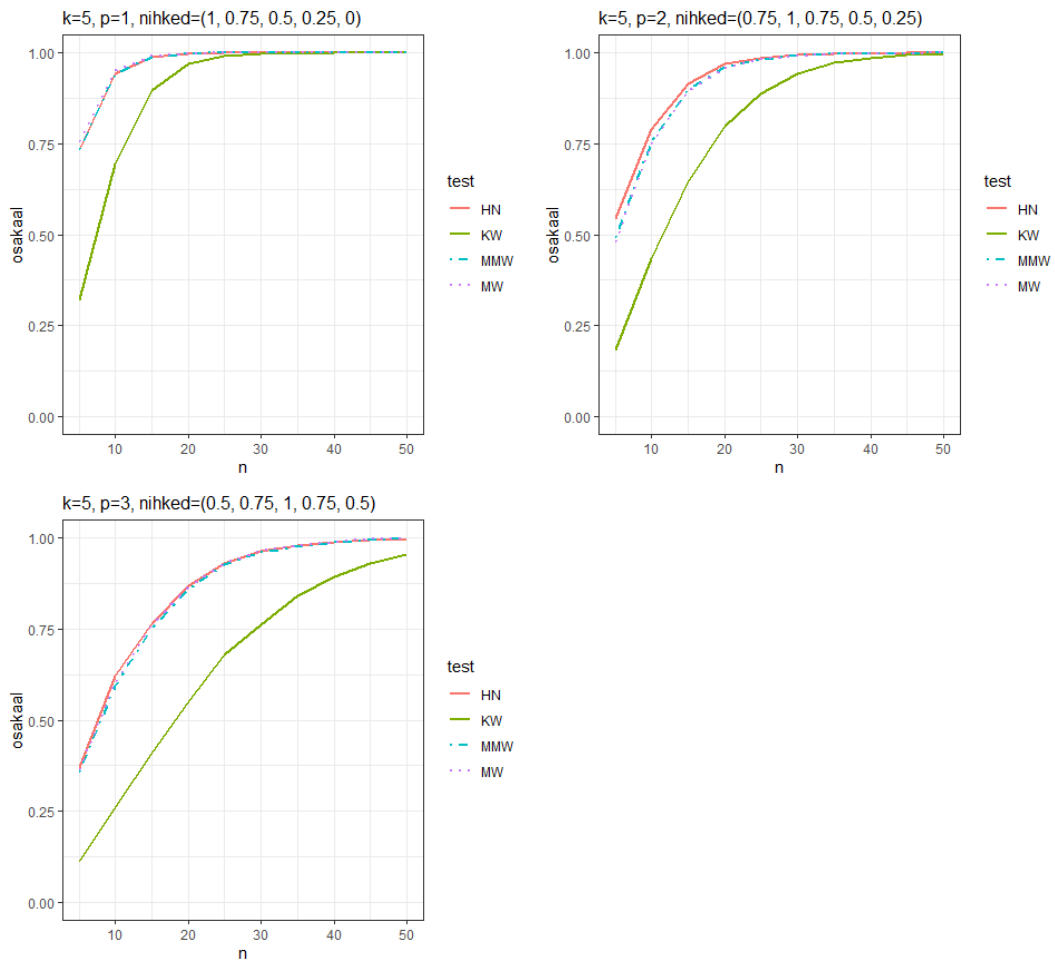
Jooniselt 10 ilmneb, et kui $k = 4$, siis $p = 1$ puhul on jällegi MW, HN ja MMW testide võimsused erinevate n väärtuste korral võrdsed. Ka $p = 2$ korral on nimetatud testid peaaegu sama võimsusega, väiksemate n väärtuste korral on HN võimsus teistest pisut suurem.

Võrreldes võimsusi eelmises alapeatükis toodud olukorraga, kus andmed olid genereeritud normaaljaotusest, siis paistab siinkohal, et eksponentjaotusega andmete korral toimub võimsuste suurenemine gruppide valimimahu n kasvades kiiremini.



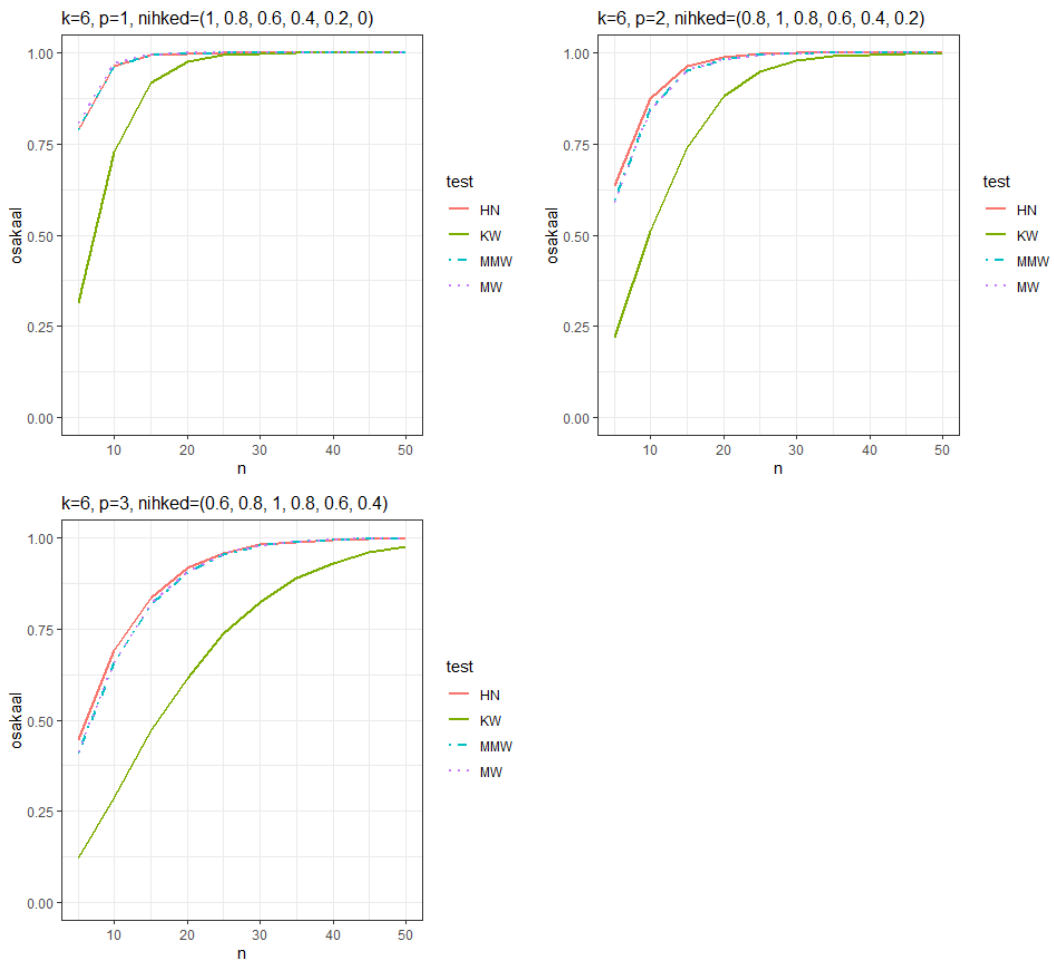
Joonis 10: Testide võimsused eksponentjaotusega andmete korral, mida on nihutatud konstandi võrra, $k = 4$

Jooniselt 11 on näha, et kui $k = 5$, siis juhtudel, kui $p = 1$ ja $p = 3$, on taas MW, HN ja MMW testid sarnase käitumisega, edestades võimsuse poolest märgatavalt KW testi vähemalt kuni selle gruppide valimimahuni n , kui kõik testid on peaaegu maksimaalse võimsusega 1. Kui $p = 2$, siis HN test on väiksemate n väärtuste juures alternatiivse hüpoteesi H_1 tuvastamisel veidi parem, kui MW ja MMW testid.



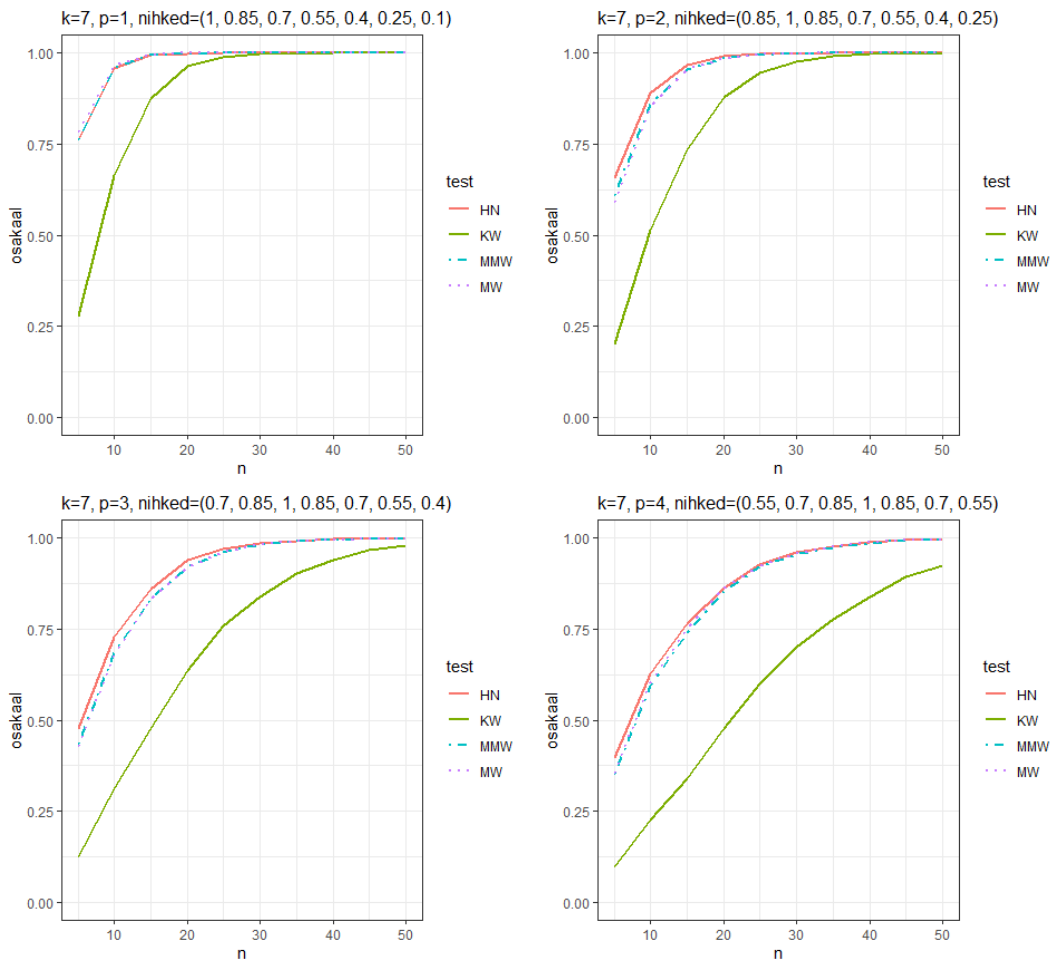
Joonis 11: Testide võimsused eksponentjaotusega andmete korral, mida on nihutatud konstandi võrra, $k = 5$

Jooniselt 12 nähtub, et kui $k = 6$, siis sarnaselt eelnevatele k väärtustele on $p = 1$ korral MW, HN ja MMW testid iga n korral võrdse võimsusega. Kui $p = 2$ või $p = 3$, siis käituvad MW ja MMW testid endiselt sama moodi, kuid HN test on neil puhkudel natukene võimsam. KW test jääb sellelgi puhul võimsuselt teistele alla.



Joonis 12: Testide võimsused eksponentjaotusega andmete korral, mida on nihutatud konstandi võrra, $k = 6$

Kui gruppe $k = 7$, siis on jooniselt 13 ähha, et iga $p \in \{1, 2, 3, 4\}$ korral käituvad MW ja MMW testid täpselt sama moodi. HN testi võimsus on $p = 1$ korral eelnevatega sama, kuid ülejäänud p valikute puhul on HN test teistest pisut suurema võimsusega. KW test on võrreldes teiste kolme testiga taas tunduvalt madalama võimsusega, mis on eriti selgelt näha sümmeetrilise kuhja olukorras ehk siis, kui $p = 4$.



Joonis 13: Testide võimsused eksponentjaotusega andmete korral, mida on nihutatud konstandi võrra, $k = 7$

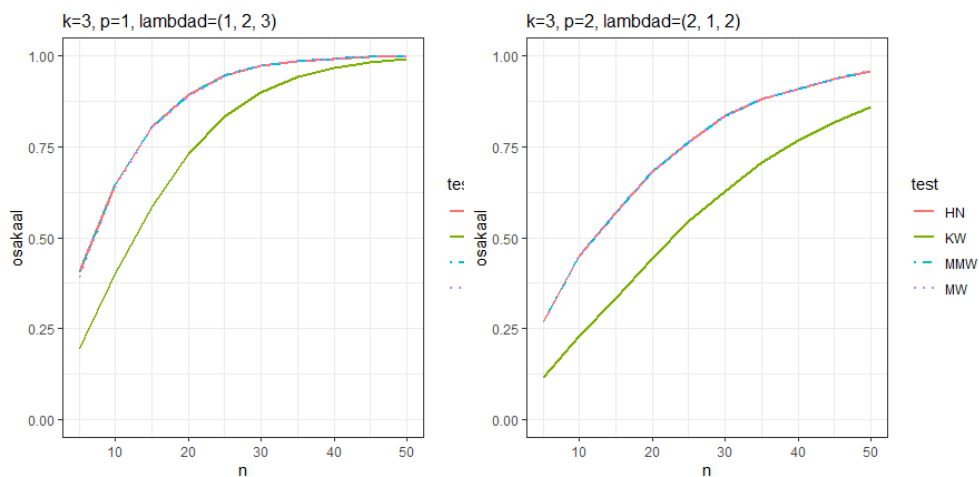
2.2.3 Eksponentjaotusega andmed erineva jaotuse parameetri korral

Selles alapeatükis pärinevad fikseeritud k korral gruppide $1, \dots, k$ vaatlused erineva parameetriga λ eksponentjaotustest. Kuna eksponentjaotuse korral on keskväärtus $\frac{1}{\lambda}$, siis andmete kuhjakujulisuse tagamiseks on kuhja tipuks valitud grupi p vaatlused genereeritud väikseima λ väärtusega eksponentjaotusest. Grupile p eelnevate ja järgnevate gruppide eksponentjaotuste parameetreid λ on suurendatud fikseeritud konstandi võrra. See aga tähendab ka seda, et mida suurem on eksponentjaotuse λ väärtus, millest on grupi vaatlused genereeritud, seda väiksem on vaatluste hajuvus selles grupis, sest eksponentjaotuse hajuvus avaldub parameetri λ kaudu $\frac{1}{\lambda^2}$.

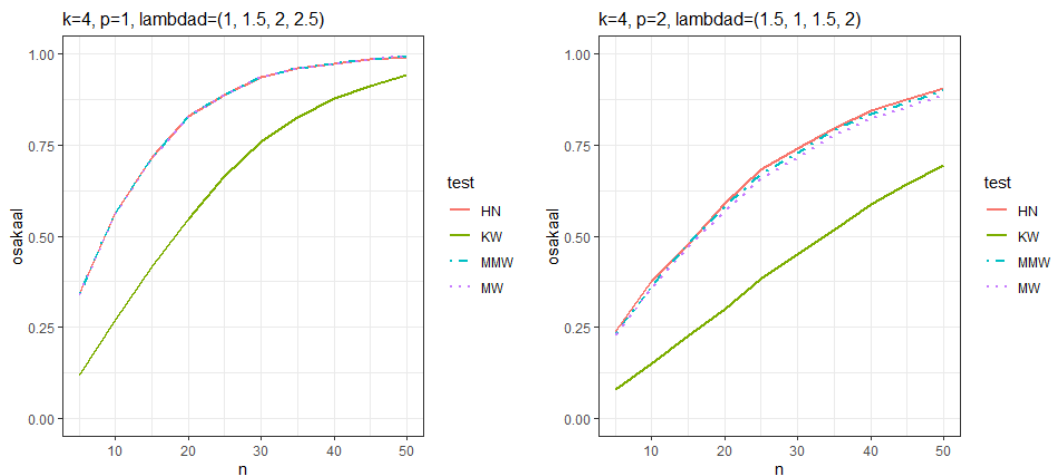
Jooniselt 14 nähtub, et kui $k = 3$, siis nii $p = 1$ kui ka $p = 2$ korral käituvad MW, HN ja MMW testid sama moodi iga n väärtuse juures. KW test on sellelgi puhul ülejäänud testidest madalama võimsusega.

Kuna $p = 1$ puhul on λ väikseim esimeses grupis, siis on selles grupis ka vaatluste hajuvus suurim ning mida kaugemal asub grupp kuhja tipust, seda väiksem on ka grupi hajuvus. Kui kuhja tipp $p = 2$, siis on suurim hajuvus keskmises grupis ning väiksem äärmistes gruppides.

Joonise 15 põhjal võib väita, et kui $k = 4$, siis $p = 1$ puhul kehtib taas, et MW, HN ja MMW testide võimsused tulevad võrdsed, edestades oluliselt KW testi võimsust. Kui $p = 2$, siis on HN testi võimsus veidi suurem MMW testi võimsusest, mis on omakorda veidi suurem MW testi võimsusest.

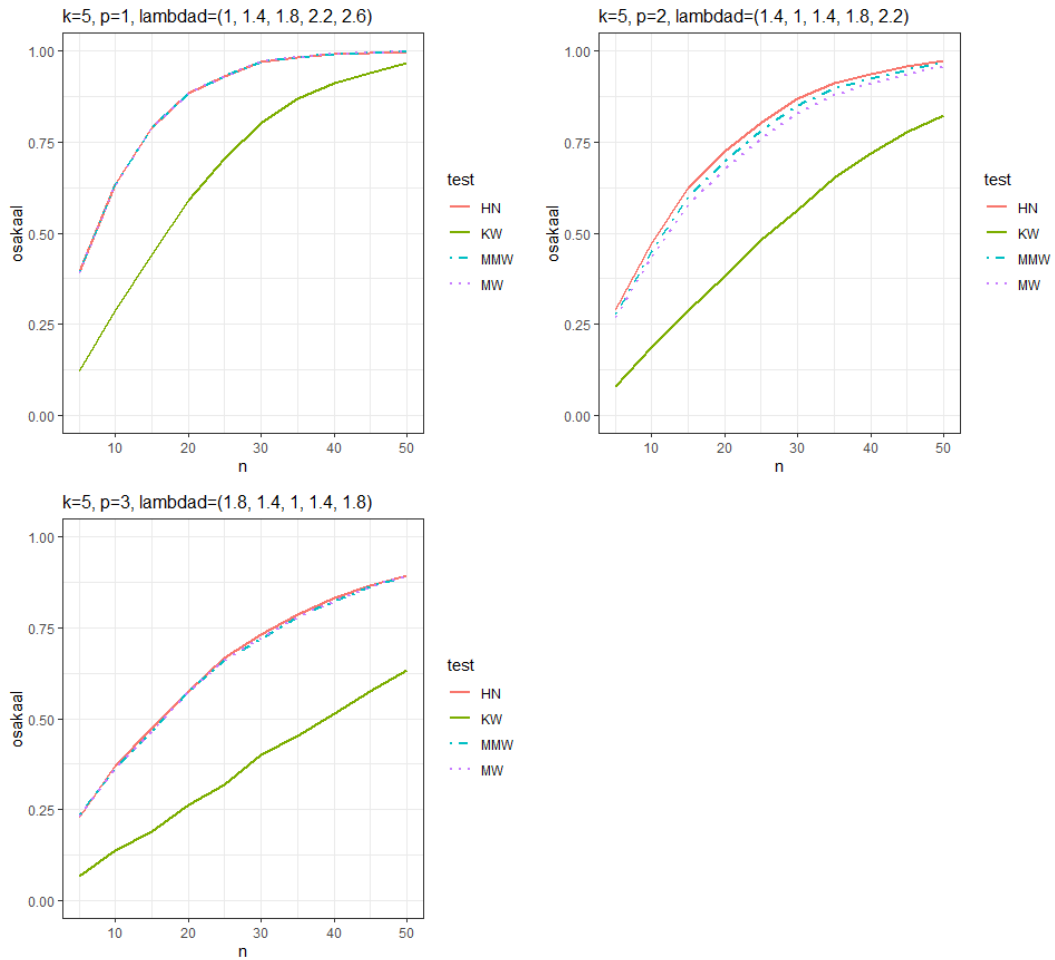


Joonis 14: Testide võimsused eksponentjaotusega andmete korral, kus on muudetud eksponentjaotuse parameetrit, $k = 3$



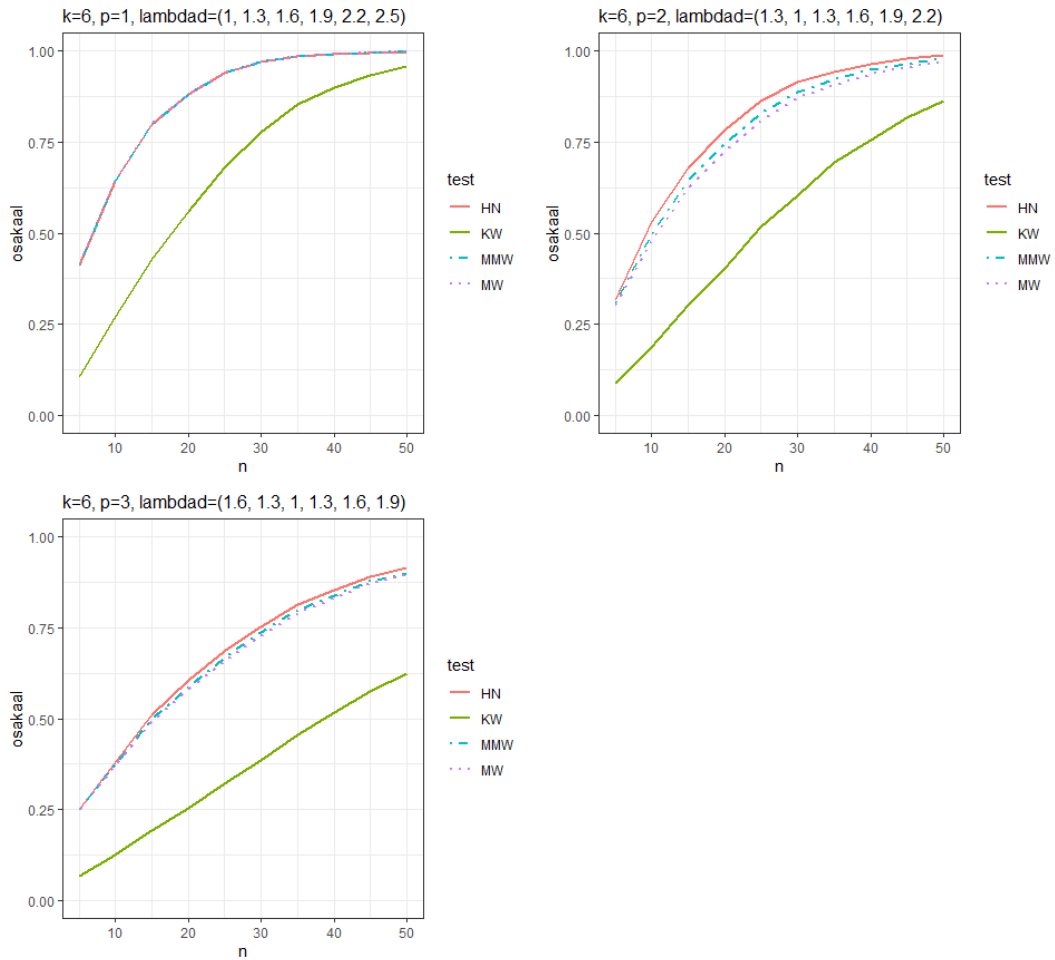
Joonis 15: Testide võimsused eksponentjaotusega andmete korral, kus on muudetud eksponentjaotuse parameetrit, $k = 4$

Jooniselt 16 on näha, et kui $k = 5$, siis $p = 1$ ning $p = 3$ on taas kolme kuhja alternatiivhüpooteeside jaoks loodud testi MW, HN ja MMW võimsused võrdsed ning KW testi võimsus neist madalam. $p = 2$ korral on HN test kõige võimsam, millele järgnevad võimsuse poolest MMW test ning MW test.



Joonis 16: Testide võimsused eksponentjaotusega andmete korral, kus on muudetud eksponentjaotuse parameetrit, $k = 5$

Joonise 17 põhjal saab öelda, et kui $k = 6$, siis $p = 1$ korral kehtib sama, nagu eelnevate k väärtuste korral. Kui $p = 2$ või $p = 3$, siis on võimsuselt suurim HN test. Teisel kohal on võimsuselt MMW test, mis napilt edestab MW testi võimsust iga n korral. KW test jääb taaskord võimsuselt ülejäänud kolmele testile alla.

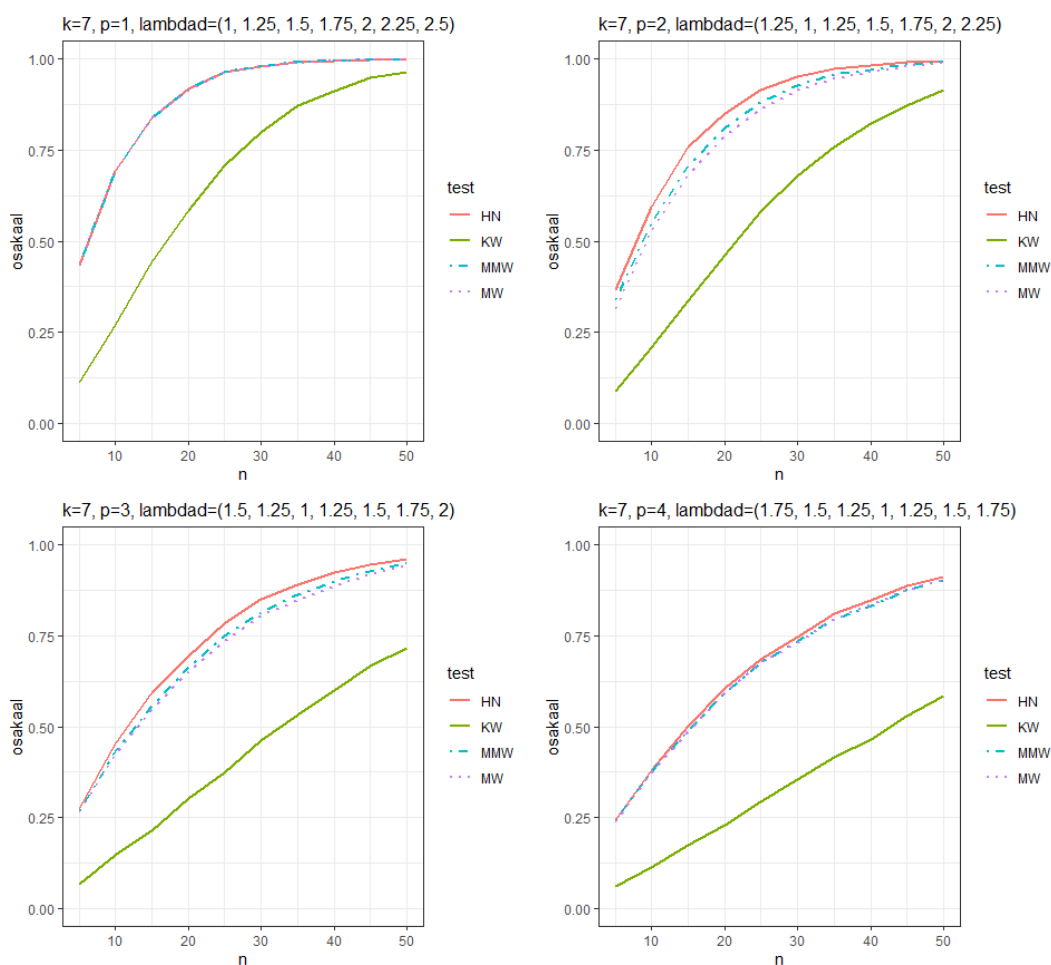


Joonis 17: Testide võimsused eksponentjaotusega andmete korral, kus on muudetud eksponentjaotuse parameetrit, $k = 6$

Jooniselt 18 on näha, et kui $k = 7$, siis $p = 1$ korral käituvad taas MW, HN ning MMW testid täpselt sama moodi. Kui $p = 4$, siis HN test on MW ja MMW testidest pisut suurema võimsusega. Kui $p = 2$ või $p = 3$, siis on HN test aga selgelt natuke suurema võimsusega kui MMW test, mis on omakorda veidi suurema võimsusega,

kui MW test. KW test on kõikide p valikute ja n väärtuste korral märgatavalt väiksema võimsusega, kui teised kolm testi.

Sarnaselt normaaljaotusega andmetega näiteks tuleb ka siin välja, et MW, HN ja MMW testid eristuvad üksteisest rohkem ebasümmeetriliste kuhjade puhul, ehk antud juhul siis, kui $p = 2$ või $p = 3$. See ei kehti aga erijuhul, kui kuhja tipp asub kõige äärmises grupis ehk $p = 1$ korral.



Joonis 18: Testide võimsused eksponentjaotusega andmete korral, kus on muudetud eksponentjaotuse parameetrit, $k = 7$

Selgus, et valitud andmete genereerimise meetodil, kus eeldused olid rikutud, tulid testide võimsused kõikidel juhtudel siiski väga sarnased olukordadele, kus andmete

genereerimisel olid kõik eeldused täidetud. Sellest ei tasu aga veel järeldada, et eelduste rikkumine ei mõjuta testide võimsusi, kuna ilmselt sõltuvad testide võimsused ka sellest, kuidas täpselt hajuvused gruppides erinevad. Tulemused võivad muutuda, kui valida teistsugune lähenemine jaotuste parameetrite valikul.

2.3 I liiki vea võrdlemine

Siinkohal uuritakse testide esimest liiki vea tegemise tõenäosuseid graafikute abil. Sarnaselt eelmise alapeatüki joonistele on ka selles alapeatükis jooniste x -teljel esitatud gruppide $n_1, \dots, n_k$ maht n ning y -teljel osakaal, kui suurel määral test 10000 genereeritud andmestikust võttis vastu alternatiivse hüpoteesi H_1 olulisuse nivool $\alpha = 0,05$. Neli uuritavat testi on esitatud eri värvi ja joonetüübiga.

Tasub silmas pidada, et gruppide $1, \dots, k$ vaatlused on antud juhul genereeritud kõik täpselt samast jaotusest ehk H_0 kehtides. Seega on oodatav tulemus, et testid võtavad alternatiivse hüpoteesi H_1 juhuslikkuse tõttu vastu umbes sama osakaaluga nagu seda on olulisuse nivoo α .

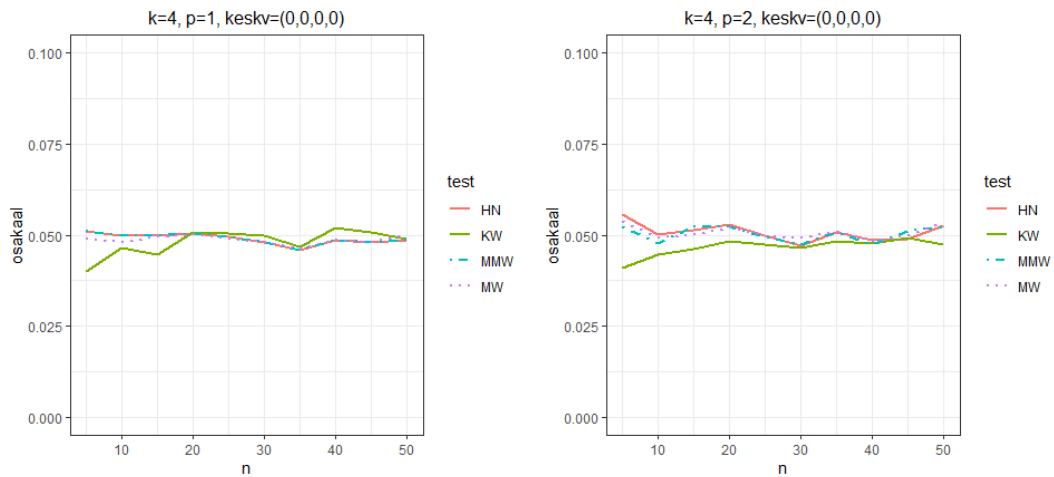
Välja on toodud esimest liiki vea tõenäosuste hinnangud nii normaaljaotusest kui ka eksponentjaotusest genereeritud andmete korral. Mõlema jaotuse kohta esitatakse joonised, kus $k = 4$, $p \in \{1, 2\}$ ning kus $k = 7$, $p \in \{1, 2, 3, 4\}$.

2.3.1 Normaaljaotusega andmed

Järgnevalt vaadatakse olukorda, kus fikseeritud k korral pärinevad kõikide gruppide $1, \dots, k$ vaatlused standardsest normaaljaotusest ehk normaaljaotusest, mille keskvärtuse parameeter $\mu = 0$ ning standardhälbe parameeter $\sigma = 1$.

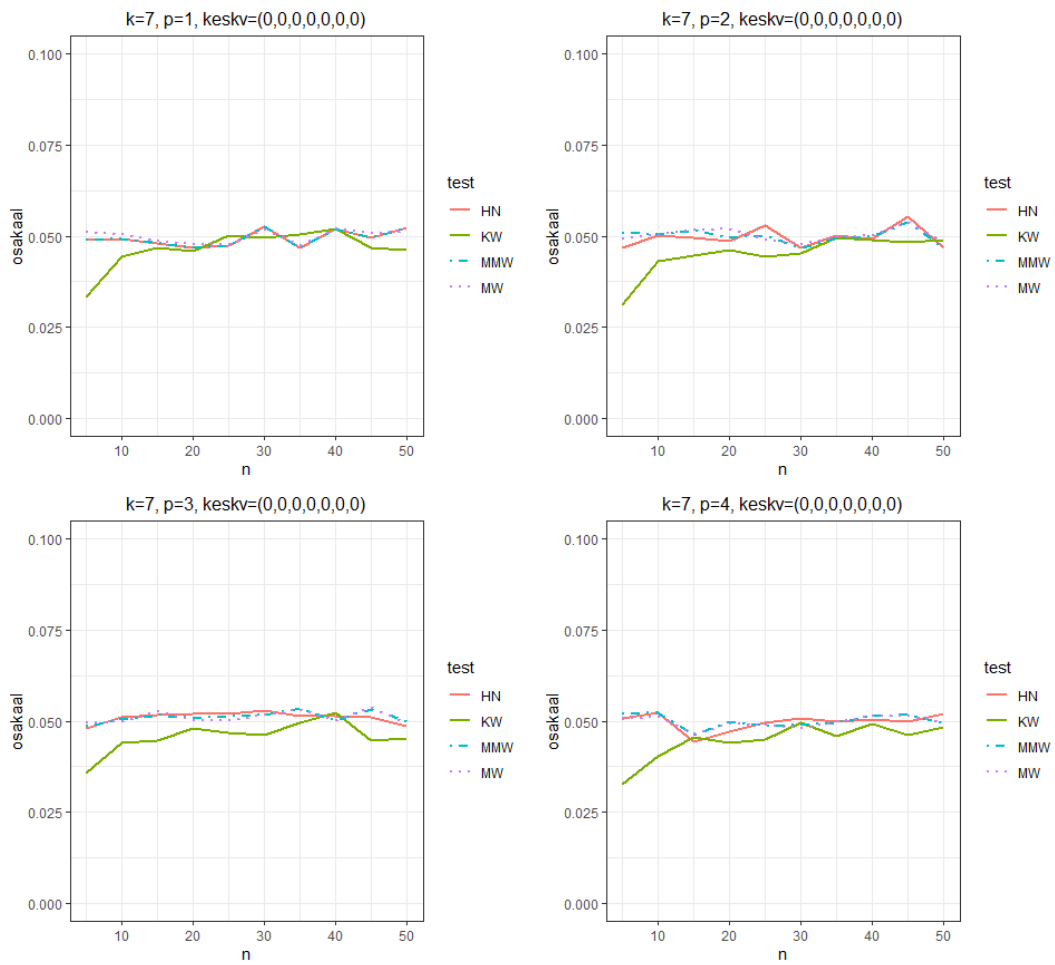
Joonise 19 põhjal võib väita, et kui $k = 4$, siis nii $p = 1$ kui ka $p = 2$ korral käituvad MW, HN ja MMW testid I liiki vea tegemise suhtes väga sarnaselt, jäädes iga n väärtuse juures olulisuse nivoo α lähedale. KW test näib käituvat veidi erinevalt, kui teised testid, mis on ka ootuspärane arvestades KW testi erinevat loomust.

Siiski on ka KW test iga uuritud n väärtuse juures üsna lähedal olulisuse nivoole α , kuid jääb mulje, et väiksemate valimimahtude juures on KW testi I liiki vea tegemise tõenäosus väiksem, kui ülejäänud testidel.



Joonis 19: Testide I liiki vea tegemise tõenäosused normaaljaotuse korral, $k = 4$

Jooniselt 20 nähtub, et kui $k = 7$, siis iga $p \in \{1, 2, 3, 4\}$ puhul on KW testi I liiki vea tegemise tõenäosus enam n väärtuste korral väiksem, kui teistel testidel. Eriti paistab see erinevus silma $n = 5$ ja $n = 10$ korral, kus KW testi I liiki vea tegemise tõenäosus olevat väiksem, kui olulisuse nivoo α . MW, HN ja MMW testid käituvad kõikidel juhtudel väga sarnaselt ning nende I liiki vea tõenäosused jäävad iga p ja n väärtuse juures väga lähedale olulisuse nivoole α .

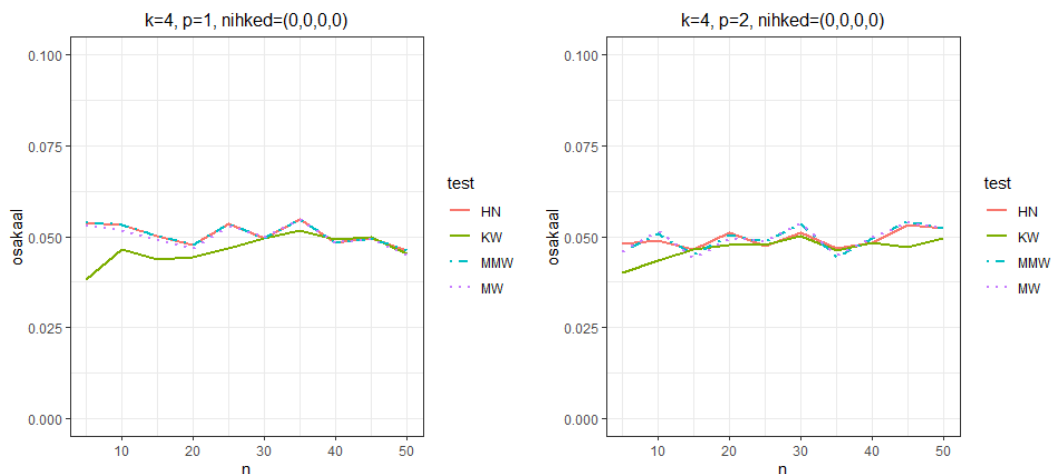


Joonis 20: Testide I liiki vea tegemise tõenäosused normaaljaotuse korral, $k = 7$

2.3.2 EkspONENTJAOTUSEGA ANDMED

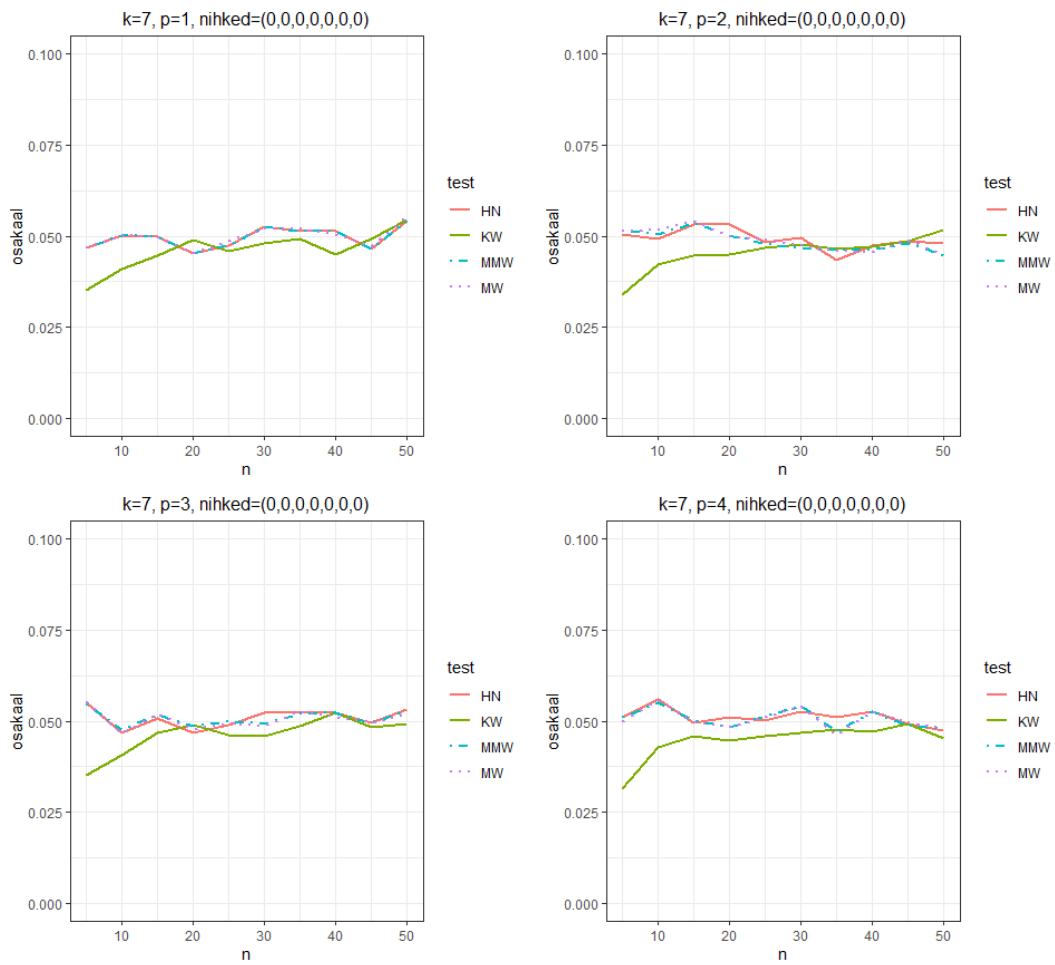
Antud alapeatükis vaadatakse olukorda, kus fikseeritud k korral pärinevad kõikide gruppide $1, \dots, k$ vaatlused eksponentjaotusest parameetriga $\lambda = 1$.

Joonise 21 põhjal saab öelda, et kui $k = 4$, siis jällegi nii $p = 1$ kui ka $p = 2$ korral käituvad MW, HN ja MMW testid väga sarnaselt, jäädes iga n juures I liiki vea tõenäosuselt olulisuse nivoo α lähedale. KW test käitub teistest testidest erinevalt ning siingi on vahe suurim väikeste n väärtuste juures.



Joonis 21: Testide I liiki vea tegemise tõenäosused eksponentjaotuse korral, $k = 4$

Jooniselt 22 on näha, et kui $k = 7$, siis iga $p \in \{1, 2, 3, 4\}$ puhul on väiksemate gruppide mahtude n juures KW test väiksema I liiki vea tegemise tõenäosusega kui teised kolm testi. MW, HN ja MMW testid käituvad endiselt üsna sarnaselt kõikide p valikute ja n väärtuste korral.



Joonis 22: Testide I liiki vea tegemise tõenäosused eksponentjaotuse korral, $k = 7$

Kokkuvõte

Käesolevas töös toodi välja Mack-Wolfe, Hettmannsperger-Nortoni ja modifitseeritud Mack-Wolfe testid, mis on mõeldud kuhja alternatiivhüpoteeside jaoks. Näidete abil tehti läbi teststatistiku arvutamine ning täpse ja asümptootilise jaotuse leidmine kõigi kolme testi korral. Selgus, et toodud näites tulid Mack-Wolfe ja modifitseeritud Mack-Wolfe testide p -väärtused täpse ja asümptootilise jaotuse kasutamise korral väga sarnased, vaatamata näite väiksest valimimahust. Hettmannsperger-Nortoni testi puhul olid p -väärtused küll mõnevõrra erinevad, kuid mõlemal juhul võeti siiski vastu alternatiivne hüpotees H_1 .

Töö praktilises osas leiti simulatsioonide abil kirjeldatud testide ning Kruskal-Wallise testi võimsused ning I liiki vea tegemise tõenäosused erinevate parameetrite korral. Vaadati kolmel erineval meetodil genereeritud andmeid ning varieeriti gruppide arvu k , kuhja tipu asukohta p ning vaatluste arvu igas grupis n .

Esitatud graafikutelt selgus, et kõikide parameetrite valiku korral käitusid kolm teooriaosas kirjeldatud testi võimsuse poolest väga sarnaselt, edestades enamike n väärtuste juures Kruskal-Wallise testi märgatavalt. Kui kuhja tipp p valiti kõige äärmiseks grupiks, siis kõnealused kolm testi käitusid kõikide uuritud variantide juures identselt.

Olukordades, kus andmete näol oli tegemist ebasümmeetrilise kuhjaga, edestas Hettmannsperger-Nortoni test võimsuselt modifitseeritud Mack-Wolfe testi, mis omakorda edestas võimsuselt Mack-Wolfe testi. Samas olid vahed kolme testi võimsustes pigem väikesed. Sümmeetriliste kuhjade korral ei tulnud nende testide võimsused üldse erinevad või olid võimsuste vahed veelgi väiksemad, kui ebasümmeetriliste kuhjade puhul.

Võrreldes omavahel testide võimsusi normaaljaotusega ja eksponentjaotusega genereeritud andmete korral, kus eeldused olid täidetud, tuli välja, et samu nihkeid kasutades saavutasid eksponentjaotuse puhul testid väiksemate n väärtuste juures

maksimaalse võimsuse. Juhul, kui kasutati eksponentjaotusega andmeid, kus konstantse hajuvuse eeldus oli rikutud, käitusid testid siiski samamoodi kui täidetud eeldustega andmete korral.

Esimest liiki vea tõenäosuse uurimisel ilmnas, et kõik neli testi käitusid nullhüpoteesi H_0 kehtides ootuspäraselt ehk alternatiivse hüpoteesi H_1 vastuvõtmise osakaal jäi kõikide parameetrite korral nii normaaljaotusega kui ka eksponentjaotusega andmete juures olulisuse nivoo 0,05 lähedale. Ainsa erisusena oli väiksemate n väärtuste juures Kruskal-Wallise testi esimest liiki vea tõenäosus mõnel määral väiksem ülejäänud kolme testi esimest liiki vea tõenäosused.

Kokkuvõtteks võib öelda, et Mack-Wolfe, Hettmannsperger-Nortoni ning modifitseeritud Mack-Wolfe testid on gruppide vahelise kuhjakujulisuse tuvastamisel tunduvalt võimsamad kui Kruskal-Wallise test. Antud teema nõuab põhjalikumat uurimist, et selgitada välja, milliste olukordade jaoks milline test kõige paremini sobib.

Kasutatud allikad

- Archambault, W. A. T., G. A. Mack ja D. A. Wolfe (1977). “k-Sample Rank Tests Using Pair-Specific Scoring Functions”. *The Canadian Journal of Statistics* 5.2, lk. 195–207. URL: <https://www.jstor.org/stable/3314778>.
- Gökpınar, E. ja F. Gökpınar (2016). “A modified Mack-Wolfe test for the umbrella alternative”. *Communications in Statistics - Theory and Methods* 45.24, lk. 7226–7241. URL: <https://doi.org/10.1080/03610926.2014.975821>.
- Hettmansperger, T. P. ja R. M. Norton (1987). “Tests for Patterned Alternatives in k-Sample Problems”. *Journal of the American Statistical Association* 82.397, lk. 292–299. URL: <https://www.jstor.org/stable/2289166>.
- Hollander, M., D. A. Wolfe ja E. Chicken (2014). *Nonparametric Statistical Methods, Third Edition*. Wiley.
- Mack, G. A. ja D. A. Wolfe (1981). “K-Sample Rank Tests for Umbrella Alternatives”. *Journal of the American Statistical Association* 76.373, lk. 175–181. URL: <https://doi.org/10.2307/2287064>.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

Lisad

A.1 Täpse jaotuse arvutamine

Siinkohal esitatakse kõik võimalikud astakute paigutused näite jaoks, kus $k = 3$, $p = 2$, $n_1 = n_2 = 2$ ning $n_3 = 1$. Samuti on iga astakute paigutuse korral välja toodud Mack-Wolfe ja Hettmannsperger-Nortoni statistikute väärtused.

Eraldi ei ole välja toodud modifitseeritud Mack-Wolfe statistiku väärtusi, sest $k = 3$ ja $p = 2$ korral langeb see kokku Mack-Wolfe statistikuga.

Tabel 6: Mack-Wolfe ja Hettmannsperger-Nortoni teststatistiku väärtused erinevate astakute paigutuste korral.

	Astakud	MW	HN		Astakud	MW	HN
1)	I II III 1 3 5 2 4	$MW_2 = 4$	$HN_2 = 0,2$	2)	I II III 1 3 4 2 5	$MW_2 = 5$	$HN_2 = 0,4$
3)	I II III 1 4 3 2 5	$MW_2 = 6$	$HN_2 = 0,6$	4)	I II III 1 2 5 3 4	$MW_2 = 4$	$HN_2 = 0$
5)	I II III 1 2 4 3 5	$MW_2 = 4$	$HN_2 = 0,2$	7)	I II III 1 4 2 3 5	$MW_2 = 6$	$HN_2 = 0,6$
7)	I II III 1 2 5 4 3	$MW_2 = 2$	$HN_2 = -0,2$	8)	I II III 1 2 3 4 5	$MW_2 = 4$	$HN_2 = 0,2$
9)	I II III 1 3 2 4 5	$MW_2 = 5$	$HN_2 = 0,4$	10)	I II III 1 2 4 5 3	$MW_2 = 2$	$HN_2 = -0,2$

	Astakud	MW	HN		Astakud	MW	HN
11)	I II III 1 2 3 5 4	$MW_2 = 3$	$HN_2 = 0$	12)	I II III 1 3 2 5 4	$MW_2 = 4$	$HN_2 = 0,2$
13)	I II III 2 1 5 3 4	$MW_2 = 2$	$HN_2 = -0,2$	14)	I II III 2 1 4 3 5	$MW_2 = 3$	$HN_2 = 0$
15)	I II III 2 4 1 3 5	$MW_2 = 6$	$HN_2 = 0,6$	16)	I II III 2 1 5 4 3	$MW_2 = 1$	$HN_2 = -0,4$
17)	I II III 2 1 3 4 5	$MW_2 = 3$	$HN_2 = 0$	18)	I II III 2 3 1 4 5	$MW_2 = 5$	$HN_2 = 0,4$
19)	I II III 2 1 4 5 3	$MW_2 = 1$	$HN_2 = -0,4$	20)	I II III 2 1 3 5 4	$MW_2 = 2$	$HN_2 = -0,2$
21)	I II III 2 3 1 5 4	$MW_2 = 4$	$HN_2 = 0,2$	22)	I II III 3 1 5 4 2	$MW_2 = 0$	$HN_2 = -0,6$
23)	I II III 3 1 2 4 5	$MW_2 = 3$	$HN_2 = 0$	24)	I II III 3 2 1 4 5	$MW_2 = 4$	$HN_2 = 0,2$
25)	I II III 3 1 4 5 2	$MW_2 = 0$	$HN_2 = -0,6$	26)	I II III 3 1 2 5 4	$MW_2 = 2$	$HN_2 = -0,2$
27)	I II III 3 2 1 5 4	$MW_2 = 4$	$HN_2 = 0$	28)	I II III 4 1 3 5 2	$MW_2 = 5$	$HN_2 = -0,6$

	Astakud	MW	HN		Astakud	MW	HN
29)	I II III 4 1 2 5 3	$MW_2 = 1$	$HN_2 = -0,4$	30)	I II III 4 2 1 5 3	$MW_2 = 2$	$HN_2 = -0,2$

A.2 Mack-Wolfe, Hettmannsperger-Nortoni ja modifitseeritud Mack-Wolfe täpne jaotus nullhüpoteesi kehtides näiteandmestiku korral

Tabelis 1 toodud näiteandmestikus on $k = 4$, $p = 3$ ja $n_1 = n_2 = n_3 = n_4 = 3$. Kokku on sellise andmestiku korral võimalikke astakute paigutusi 369600. Allpool olevates tabelites on esitatud statistikute kõikvõimalikud väärtused ning nende esinemiste sagedused. Et leida statistiku konkreetse väärtuse saamise tõenäosust nullhüpoteesi kehtides, tuleb väärtuse esinemise sagedus labi jagada kõikvõimalike astakute paigutuste arvuga.

Tabel 7: Mack-Wolfe statistiku väärtuste esinemiste sagedused nullhüpoteesi kehtides.

<i>MW</i>	sagedus	<i>MW</i>	sagedus	<i>MW</i>	sagedus	<i>MW</i>	sagedus
0	84	10	10584	20	22848	30	3444
1	168	11	12936	21	21588	31	2268
2	420	12	15372	22	19740	32	1428
3	840	13	17640	23	17640	33	840
4	1428	14	19740	24	15372	34	420
5	2268	15	21588	25	12936	35	168
6	3444	16	22848	26	10584	36	84
7	4788	17	23688	27	8484		
8	6468	18	24024	28	6468		
9	8484	19	23688	29	4788		

Tabel 8: Hettmannsperger-Nortoni statistiku väärtuste esinemiste sagedused nullhüpoteesi kehtides.

HN	sagedus	HN	sagedus	HN	sagedus	HN	sagedus
$-\frac{27}{12}$	20	$-\frac{13}{12}$	6240	$\frac{1}{12}$	16400	$\frac{15}{12}$	4620
$-\frac{26}{12}$	40	-1	6800	$\frac{2}{12}$	15400	$\frac{16}{12}$	3680
$-\frac{25}{12}$	100	$-\frac{11}{12}$	8340	$\frac{3}{12}$	15320	$\frac{17}{12}$	3100
-2	200	$-\frac{10}{12}$	9200	$\frac{4}{12}$	14280	$\frac{18}{12}$	2360
$-\frac{23}{12}$	360	$-\frac{9}{12}$	10580	$\frac{5}{12}$	14020	$\frac{19}{12}$	1880
$-\frac{22}{12}$	600	$-\frac{8}{12}$	11200	$\frac{6}{12}$	12920	$\frac{20}{12}$	1320
$-\frac{21}{12}$	980	$-\frac{7}{12}$	12440	$\frac{7}{12}$	12440	$\frac{21}{12}$	980
$-\frac{20}{12}$	1320	$-\frac{6}{12}$	12920	$\frac{8}{12}$	11200	$\frac{22}{12}$	600
$-\frac{19}{12}$	1880	$-\frac{5}{12}$	14020	$\frac{9}{12}$	10580	$\frac{23}{12}$	360
$-\frac{18}{12}$	2360	$-\frac{4}{12}$	14280	$\frac{10}{12}$	9200	2	200
$-\frac{17}{12}$	3100	$-\frac{3}{12}$	15320	$\frac{11}{12}$	8340	$\frac{25}{12}$	100
$-\frac{16}{12}$	3680	$-\frac{2}{12}$	15400	1	6800	$\frac{26}{12}$	40
$-\frac{15}{12}$	4620	$-\frac{1}{12}$	16400	$\frac{13}{12}$	6240	$\frac{27}{12}$	20
$-\frac{14}{12}$	5240	0	14320	$\frac{14}{12}$	5240		

Tabel 9: Modifitseeritud Mack-Wolfe statistiku väärtuste esinemiste sagedused nullhüpoteesi kehtides.

<i>MMW</i>	sagedus	<i>MMW</i>	sagedus	<i>MMW</i>	sagedus	<i>MMW</i>	sagedus
0	84	12	8363	24	17100	36	5090
1	168	13	9654	25	16665	37	4147
2	420	14	10956	26	16114	38	3364
3	840	15	12302	27	15421	39	2624
4	1307	16	13440	28	14531	40	1931
5	1931	17	14531	29	13440	41	1307
6	2624	18	15421	30	12302	42	840
7	3364	19	16114	31	10956	43	420
8	4147	20	16665	32	9654	44	168
9	5090	21	17100	33	8363	45	84
10	5948	22	17270	34	7061		
11	7061	23	17270	35	5948		

A.3 R kood andmete genereerimiseks

```
### Normaaljaotusega andmed ###
```

```
normaal_andmed <- function(k, n, keskval){  
  vaatlused <- c()  
  grupid <- c()  
  for (i in 1:k) {  
    vaatlused <- c(vaatlused, rnorm(n, mean = keskval[i]))  
    grupid <- c(grupid, rep(i, times = n))  
  }  
  return(data.frame(vaatlused, grupid))  
}
```

```
### Eksponentjaotusega andmed ###
```

```
# Sama lambda, nihutatud
```

```
eksp_andmed1 <- function(k, n, lambda, nihked){  
  vaatlused <- c()  
  grupid <- c()  
  for (i in 1:k) {  
    vaatlused <- c(vaatlused, rexp(n, rate = lambda) + nihked[i])  
    grupid <- c(grupid, rep(i, times = n))  
  }  
  return(data.frame(vaatlused, grupid))  
}
```

```
#Erinev lambda
```

```
eksp_andmed2 <- function(k, n, lambdad){  
  vaatlused <- c()  
  grupid <- c()  
  for (i in 1:k) {  
    vaatlused <- c(vaatlused, rexp(n, rate = lambdad[i]))  
  }
```

```
      grupid <- c(grupid, rep(i, times = n))
    }
    return(data.frame(vaatlused, grupid))
  }
}
```


A.4 R kood statistikute väärtuste ja testide p-vääruste leidmiseks

```
# Funktsioon Mann-Whitney statistiku leidmiseks
# kahe grupi vaatluste vahel
MannWhitney <- function(grupp1, grupp2){
  tulemus <- 0
  for (vaatlus1 in grupp1) {
    for (vaatlus2 in grupp2){
      tulemus <- tulemus + ifelse(vaatlus1 < vaatlus2, 1, 0)
    }
  }
  return(tulemus)
}

### Mack-Wolfe ###

# Abifunktsioon: leiab Mack Wolfe statistiku vaartuse
MW_stat <- function(vaatlused, grupid, p) {
  k <- as.numeric(max(grupid))
  MW <- 0
  if (p != 1){
    for (i in 1:(p-1)){
      for (j in (i+1):p){
        MW <- MW +
          MannWhitney(vaatlused[grupid == i],
                      vaatlused[grupid == j])
      }
    }
  }
  if (p != k){
    for (i in p:(k-1)){
      for (j in (i+1):k){

```

```

        MW <- MW +
            MannWhitney(vaatlused[grupid == j],
                        vaatlused[grupid == i])
    }
}
}
return(MW)
}

# Funktsioon modifitseeritud Mack Wolfe statistiku vaartuse
# leidmiseks ja standardiseerimiseks ning testi tegemiseks.
# Valjastab standardiseeritud statistiku vaartuse
# ning testi p-vaartuse.
MW_stand <- function(vaatlused, grupid, p){
    MW <- MW_stat(vaatlused, grupid, p)
    # Valimimahud
    N_1 <- length(vaatlused[grupid <= p])
    N_2 <- length(vaatlused[grupid >= p])
    N <- length(vaatlused)
    n_p <- length(vaatlused[grupid == p])
    nj_ruut_summa <- 0
    for (grupp in unique(grupid)) {
        nj <- length(vaatlused[grupid == as.numeric(grupp)])
        nj_ruut_summa <- nj_ruut_summa + nj^2
    }
    nj_ruut_summa_2 <- 0
    for (grupp in unique(grupid)) {
        nj <- length(vaatlused[grupid == as.numeric(grupp)])
        nj_ruut_summa_2 <- nj_ruut_summa_2 + nj^2*(2*nj+3)
    }
    # Keskvaaartus
    E_MW <- (N_1^2 + N_2^2 - nj_ruut_summa - n_p^2)/4
    # Dispersioon

```

```

D_MW <- 1/72*(2*(N_1^3 + N_2^3) + 3*(N_1^2 + N_2^2)
        - nj_ruut_summa_2 - n_p^2*(2*n_p + 3)
        + 12*n_p*N_1*N_2 - 12*n_p^2*N)
statistik <- (MW-E_MW)/sqrt(D_MW)
p_vaartus <- 1 - pnorm(statistik)
return(c(statistik, p_vaartus))
}

### Hettmansperger-Norton ###

# Funktsioon Hettmannsperger-Nortoni statistiku vaartuse
# leidmiseks ja standardiseerimiseks ning testi tegemiseks.
# Valjastab standardiseeritud statistiku vaartuse
# ning testi p-vaartuse.
HN_stand <- function(vaatlused, grupid, p){
  k <- length(unique(grupid)) # gruppide arv
  N <- length(vaatlused) # vaatluste arv
  astakud <- rank(vaatlused)
  keskastakud <- rep(NA, k)
  c <- rep(NA, k)
  lambda <- rep(NA, k)
  for (i in 1:k) {
    keskastakud[i] <-
      sum(astakud[grupid == i])/sum(grupid == i)
    c[i] <- ifelse(i <= p, i, 2*p-i)
    lambda[i] <- sum(grupid == i)/N
  }
  c_w <- sum(c*lambda)
  HN <- round(sum(lambda*(c-c_w)*keskastakud), 10)
  # Dispersioon
  D_HN <- (N+1)/12*sum(lambda*(c-c_w)^2)
  statistik <- HN/sqrt(D_HN)

```

```

    p_vaartus <- 1 - pnorm(statistik)
    return(c(statistik, p_vaartus))
}

### Modifitseeritud Mack-Wolfe test ###

# Abifunktsioon: leiab modifitseeritud
# Mack Wolfe statistiku vaartuse
MMW_stat <- function(vaatlused, grupid, p) {
  k <- as.numeric(max(grupid))
  MMW <- 0
  if (p != 1){
    for (i in 1:(p-1)){
      for (j in (i+1):p){
        MMW <- MMW + (j-i)*MannWhitney(
          vaatlused[grupid == i], vaatlused[grupid == j])
      }
    }
  }
  if (p != k){
    for (i in p:(k-1)){
      for (j in (i+1):k){
        MMW <- MMW + (j-i)*MannWhitney(
          vaatlused[grupid == j], vaatlused[grupid == i])
      }
    }
  }
  return(MMW)
}

# Funktsioon modifitseeritud Mack Wolfe statistiku vaartuse
# leidmiseks ja standardiseerimiseks ning testi tegemiseks.
# Valjastab standardiseeritud statistiku vaartuse

```

```

# ning testi p-vaartuse.
MMW_stand <- function(vaatlused, grupid, p){
  # NB! EELDAB, ET KOIK GRUPID ON VORDSE SUURUSEGA
  MMW <- MMW_stat(vaatlused, grupid, p)
  n <- length(vaatlused[grupid == p]) # yhe grupi valimimahd
  k <- length(unique(grupid)) # gruppide arv
  # Keskvaartus
  E_MMW <- n^2/2*(choose(p+1, 3) + choose(k-p+2, 3))
  # Dispersioon
  D_MMW <- (n^2*p^2*(p^2-1)*(n*p+1) + n^2*(k-p+1)^2*
            ((k-p+1)^2-1)*(n*(k-p+1)+1))/144 +
            (n^3*p*(p-1)*(k-p)*(k-p+1))/24
  statistik <- (MMW-E_MMW)/sqrt(D_MMW)
  p_vaartus <- 1 - pnorm(statistik)
  return(c(statistik, p_vaartus))
}

```

A.5 R kood testide tegemiseks

Siinkohal on esitatud vaid normaaljaotusega andmetega testide tegemise funktsioon. Teiste kahe jaotusega testide tegemiseks piisab vaid sellest, kui muuta reas `andmed <- normaal_andmed(k, n, keskval)` andmete genereerimise funktsiooni ning anda sellele funktsioonile sisendina nõutavad parameetrid. Sarnane muudatus tuleb teha ka koodi esimeses reas, kus funktsiooni sisendite nimekiri tuleb muuta vastavusse andmete genereerimise funktsiooniga, mida soovitakse kasutada.

```
### Normaaljaotusega andmed ###
normaal_test <- function(k, p, n, keskval, kordusi){
  tulemused_KW <- rep(NA, kordusi)
  tulemused_MW <- rep(NA, kordusi)
  tulemused_MMW <- rep(NA, kordusi)
  tulemused_HN <- rep(NA, kordusi)
  for (i in 1:kordusi) {
    andmed <- normaal_andmed(k, n, keskval) # Seda muuta
    tulemused_KW[i] <- kruskal.test(
      vaatlused ~ grupid, data = andmed)$p.value < 0.05
    tulemused_MW[i] <- MW_stand(
      andmed$vaatlused, andmed$grupid, p)[2] < 0.05
    tulemused_MMW[i] <- MMW_stand(
      andmed$vaatlused, andmed$grupid, p)[2] < 0.05
    tulemused_HN[i] <- HN_stand(
      andmed$vaatlused, andmed$grupid, p)[2] < 0.05
  }
  test <- c("KW", "MW", "MMW", "HN")
  osakaal <- c(sum(tulemused_KW), sum(tulemused_MW),
               sum(tulemused_MMW), sum(tulemused_HN))/kordusi
  print(n)
  print(Sys.time())
  return(data.frame(test, osakaal))
}
```

A.6 R kood jooniste väljastamiseks

Siinkohal esitatakse vaid kood normaaljaotusega andmetega joonise väljastamiseks. Teiste kahe jaotuse jooniste tekitamiseks tuleb vaid muuta rida `fikseeritud_n` `<- normaal_test(k, p, n, kesk, kordusi)`, vahetades sealne testide tegemise funktsioon ära soovitud jaotuse testide tegemise funktsiooniga. Lisaks tuleb muuta ka funktsiooni esimest rida, et sealsed sisendparameetrid oleksid vastavuses eelnevalt muudetud testide tegemise funktsiooni sisendparameetritega.

```
### Normaaljaotusega andmed ###
normaal_joonis <- function(k, p, n_algus, n_lopp,
                           n_kaupa, kesk, kordusi, pealkiri){
  tulemus <- data.frame(test = character(),
                        osakaal = numeric(), n = numeric())
  for(n in seq(
    from = n_algus, to = n_lopp, by = n_kaupa)){
    fikseeritud_n <- normaal_test(
      k, p, n, kesk, kordusi)
    fikseeritud_n$n <- n
    tulemus <- rbind(tulemus, fikseeritud_n)
  }

  ggplot(data = tulemus,
    mapping = aes(x = n,
      y = osakaal, col = test, lty = test)) +
    geom_line(size = 1) +
    scale_linetype_manual(
      values=c("solid","solid", "dotdash", "dotted")) +
    ylim(0, 1) +
    ggtitle(pealkiri) +
    theme_bw() + theme(plot.title = element_text(size=12))
}
```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Raul Niit,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Mittepameetrilised testid kuhja alternatiivhüpoteesidele, 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.

Raul Niit

18.05.2021