

TARTU ÜLIKOOL  
MATEMAATIKA-INFORMAATIKATEADUSKOND  
MATEMAATILISE STATISTIKA INSTITUUT

Edwart Ždanovitš

Barnardi täpse testi võrdlus Fisheri täpse testiga

Bakalaureusetöö

Juhendaja:

Märt Möls

TARTU

2013

## Sisukord

|   |    |
|---|----|
| Sissejuhatus.....   | 3  |
| Barnardi täpne test .....   | 5  |
| Waldi teststatistik.....  | 5  |
| Barnardi testi tööpõhimõte.....   | 5  |
| Fisheri täpne test .....  | 11 |
| Hüpergeomeetriline jaotus .....   | 11 |
| Fisheri täpse testi tööpõhimõte .....                                       | 11 |
| Barnardi testi ja Fisheri testi võimsus .....                               | 13 |
| Testi diskreetsus.....  | 13 |
| Suurematele kui 2 x 2 sagedustabelitele rakendatud testide diskreetus ..... | 16 |
| Tõepärasuhte test .....   | 17 |
| Võimsusfunktsioon .....   | 18 |
| Probleemid Barnardi p-väärtuse leidmisel.....                               | 25 |
| Kokkuvõte.....  | 27 |
| Barnard's and Fisher's exact test comparison .....                          | 28 |
| Summary .....   | 28 |
| Kirjanduse loetelu .....  | 29 |
| LISA 1.....   | 30 |

## Sissejuhatus

Käesoleva töö eesmärgiks on võrrelda kahte erinevat testi, Barnardi ja Fisheri täpset testi. Peamine erinevus testide vahel seisneb tingimuslikkuses. Seda kas sagedustabelite kohta käivate hüpoteeside testimiseks kasutada pigem tingimuslikku või tingimusteta testi, on olnud kõneaineks juba üle 50 aasta. Tingimuslikkusega kaasnevad testi omaduste erinevused nagu näiteks testi võimsus ning tõenäosus teha I liiki viga. Üheks erinevuste põhjuseks võib lugeda  $2 \times 2$  sagedustabelite põhjal leitud teststatistikute liigset diskreetsust. Kogu töö vältel on uurimise all peamiselt  $2 \times 2$  mõõtmelised sagedustabelid. Testide diskreetsuse kirjeldamisel puututakse kokku vähesel määral ka  $2 \times m$  mõõtmelisi sagedustabeleid.

Käsitletavad  $2 \times 2$  mõõtmelised sagedustabelid on üles ehitatud binaarsetele andmetele, ehk eristatakse mingi sündmuse esinemiste/mitte esinemiste hulka kogu mõõtmistulemuste hulgast.  $2 \times 2$  sagedustabelid leiavad kasutust paljudel pragmaatilistel aladel, kus nõutakse testi võimalikult suurt võimsust. Üheks selliseks valdkonnaks on meditsiin, kus mõõdetakse mingi ravi või vaktsiini ja farmakoloogiliselt toimetu ravimi manustamise tagajärjel terveks saanute või nakatunute hulka kõigi katses osalenute seas. Meditsiinilised eksperimendid võivad olla kulukad või muude ressursside poolt piiratud, mistõttu valimid on väiksed. Sarnastes katsetes võib testist tuleneval otsusel olla määrav tähtsus inimese paranemisel või surmast pääsemisel, mis seda enam suurendab nõudlust õigema testi valiku järele.

Töö esimesed kaks peatükki koos alapeatükkidega annavad ülevaate nii Barnardi testi, kui ka Fisheri testi definitsioonist ning tööpõhimõttest. Seletatakse lahti Waldi teststatistik ning tema kasutuse eesmärk Barnardi testi rakendamisel. Kasutatakse ka näiteandmestikku arvutuskäikude seletamiseks.

Kolmas peatükk on jaotatud neljaks alampeatükis, millest esimeses seletatakse lahti testi diskreetsus ning seda suurendavad ja vähendavad tegurid. Põhjendatakse, miks on Barnardi täpne test mingites kindlates situatsioonides võimsam Fisheri täpsest testist ning miks Barnardi ehk tingimusteta testi võimsuse eelis tabeli mõõtmete ja valimi mahu suurenedes kaob. Parema võrdlusmomendi saamiseks kaasatakse vaadeldavasse peatükki ka tõepärasuhte test, mis erinevalt eelnevalt kirjeldatutest ei ole täpne test. Viimases osas kirjeldatakse võimsusfunktsiooni definitsiooni ning leitakse nii Barnardi, Fisheri, kui ka tõepärasuhte testi võimsus sõltuvalt üldkogumis eksisteerivast parameerist ning sagedustabelile seatud tingimustest. Võimsuste erinevused on kirjeldatud graafikute alusel.

Viimases peatükis rakendatakse nii Barnardi kui ka Fisheri testi järjest suuremate valimite sagedustabelitele ning mõõdetakse testi läbiviimiseks kuluvat aega.

## Barnardi täpne test

Barnardi täpset testi kasutatakse hüpoteeside kontrollimiseks väikeste valimimahtudega tabelite põhjal. Testi kirjelduse avaldas George Alfred Barnard 1945. aastal, kes väitis, et tema välja pakutud test rakendatuna 2 x 2 sagedustabelil on võimsam kui Fisheri täpne test (Wikipedia, 2013a). Barnardi testi ja Fisheri testi suurim erinevus seisneb tingimuslikkuses, kus esimene neist on tingimusteta ning teine tingimuslik test. Testide tingimuslikkus seisneb testi läbiviimisel uuritavate juhuslike tabelite kaasamise ja mittekaasamise otsustes, mille täpsem seletus antakse järgnevatel peatükkides. Barnardi testi puhul vabanetakse üldkogumis eksisteeriva sündmuse toimumise tõenäosuse hindamisest teda kirjeldava parameetri maksimeerimise teel ehk leitakse testi seisukohalt suurimat p-väärtust väljastav parameeter, mis võib üldkogumis aset leida.

## Waldi teststatistik

Uurimustöös kasutatakse Barnardi testi teoreetilise tõlgenduse jaoks Waldi teststatistikut. Abraham Waldi järgi nime saanud Waldi teststatistik on laialdaselt rakendust leidnud ning eelkõige kasutatakse seda hinnatud parameetri täpsuse testimiseks. Waldi teststatistik avaldub kujul:

$$W = \frac{(\hat{\theta} - \theta_0)}{se(\hat{\theta})}, \quad (1)$$

kus  $\hat{\theta}$  on uuritava parameetri  $\theta$  hinnang,  $\theta_0$  on üldkogumi õige väärtus või eelnevalt paika pandud parameeter ning  $se(\hat{\theta})$  on hinnatud parameetri standardhälve. (Wikipedia, 2013b).

## Barnardi testi tööpõhimõte

Käesolev alampeatükk põhineb Mehta ja Senchaudhuri (2003) artiklil. Barnardi testi tööpõhimõte seletatakse lahti 2 x 2 sagedustabeli näitel, mis kirjeldab vaktsiini mõju patsientide haigestumisele. Vaatluse all on  $N$  objektiline valim. Nendest osad ehk  $n_{+1}$  suvaliselt valitud patsiendid vaktsineeritakse vaktsiiniga ning ülejäänud  $n_{+2}$  farmakoloogiliselt toimetu ravimiga ehk platseeboga. Peale patsientidele ravimite manustamist uuritakse nende immuunsust neid vaktsineeritud haiguse vastu. Olgu vaktsiini saanud ning haigestunud, vaktsiini saanud ja mitte haigestunud, platseebot manustanud ning haigestunud ja platseebot manustanud ning mitte haigestunute arvud tabelis tähistatud

vastavalt  $x_{11}$ ,  $x_{21}$ ,  $x_{12}$ , ja  $x_{22}$  (vaata tabel 1). Tähistagu  $\frac{x_{11}}{x_{11}+x_{21}} = \hat{p}_v$  haigestunute osakaalu eksperiment ehk vaktsineeritute grupis ning  $\frac{x_{12}}{x_{12}+x_{22}} = \hat{p}_k$  haigestunute osakaalu kontroll ehk platseebot saanute grupis. Vaadeldavas katses nullhüpooteesi kehtides on nii vaktsineeritute grupis, kui ka kontroll grupis osalenute tõenäosused haigestuda võrdsed ehk

$$H_0: p_v = p_k, \quad (2)$$

kus  $p_v$  tähistab tõenäosust haigestuda peale vaktsiini manustamist ning  $p_k$  tähistab tõenäosust nakatuda ilma vaktsiin saamata.

Tabel 1. Haigestumise ja vaktsineerimise seost iseloomustav sagedustabel

| Haigestunud | Manustatud ravim |           | Kokku                 |
|-------------|------------------|-----------|-----------------------|
|             | Vaktsiin         | Platseebo |                       |
| Jah         | $x_{11}$         | $x_{12}$  | $n_{1+}$              |
| Ei          | $x_{21}$         | $x_{22}$  | $n_{2+}$              |
| Kokku       | $n_{+1}$         | $n_{+2}$  | $n_{+1} + n_{+2} = N$ |

Tähistagu  $X_0$  katse tulemusena saadud sagedustabelit ning  $X$  kõikvõimalikke sündmuste toimumiste kombinatsioonidega  $2 \times 2$  tabelite hulka selliselt, et  $n_{+1}$  ja  $n_{+2}$  konstantsed. Tabelis olev väärtus  $x_{21}$  avaldub kõigi vaktsineeritute ja vaktsineeritute, kuid mitte haigestunute vahena ehk  $x_{21} = n_{+1} - x_{11}$ , seega tunnus vaktsiin on binaarne ning vaktsineeritute grupi väärtused on binoomjaotusega (vaata tabel 1). Analoogne tulemus kehtib ka kontrollgrupis. Järelikult nii vaktsiini saanute grupis, kui ka kontrollgrupis on juhusliku haigestunute ja mittehaigestunute suhte tekkimise tõenäosus arvatav binoomjaotuse tõenäosusvalemist, kus katsete arv esimeses grupis  $n_{+1}$  ja teises grupis  $n_{+2}$  ning sündmuste toimumiste ehk antud näite puhul haigestumiste arv vastavalt  $x_{11}$  ja  $x_{12}$ . Kuna juhuslik võimalik  $2 \times 2$  sagedustabel on vaktsiini saanute grupi jaotuse ja kontrollgrupi jaotuse koosjuhtumise ehk koostoimimise tõenäosus ning gruppide jaotused on üksteisest sõltumatud, siis tõenäosus vaadeldava tabeli tekkimiseks avaldub kummagi grupi jaotuse tekkimise tõenäosuste omavahelisel korrutamisel. Kui eeldada, et nullhüpootees kehtib, tõenäosused  $p_v$  ja  $p_k$  on võrdsed ehk  $p_v = p_k = p$ , siis väärtustega  $x_{11}$  ja  $x_{12}$  sagedustabeli tekkimise tõenäosus on arvatav kahekordse binoomjaotuse tõenäosusvalemist:

$$P(X_0|p) = \binom{n+1}{x_{11}} \binom{n+2}{x_{12}} p^{(x_{11}+x_{12})} (1-p)^{(N-x_{11}-x_{12})}. \quad (3)$$

Juhul, kui vaatluse all on  $2 \times m$  mõõtmeline ehk  $m$  grupist koosnev sagedustabel, siis nullhüpoteesi  $p_1 = p_2 = \dots = p_m = p$  kehtides saab sündmuste toimumise väärtustega  $x_{11}, x_{12}, \dots, x_{1m}$  sagedustabeli tekkimiste tõenäosust arvutada  $m$ -kordsest binoomjaotuse tõenäosusvalemist:

$$P(X_0|p) = \prod_{i=1}^m \binom{n+i}{x_{1i}} p^{\sum_{i=1}^m x_{1i}} (1-p)^{(N-\sum_{i=1}^m x_{1i})}, \quad (3')$$

kus  $x_{1i}$  tähistab sündmuste toimumiste arv grupis  $i$  ning  $n_i$  tähistab grupi  $i$  suurust.

Eelneva mõttekäigu paremaks mõistmiseks vaadatakse näidet, kus vaatluse all on valim 30 patsiendiga, kellest 15-le on manustatud vaktsiini ning ülejäänutele platseebot. Katse lõpuks oli haigusse nakatunuid vaktsiini saanute hulgas 7 ning platseebo ehk kontrollgrupis 12 (vaata tabel 2). Sellisel juhul ühepoolne täpne  $p$ -väärtus avaldub sama ekstreemse või veel ekstreemsemate tabelite kui  $X_0$  tekkimise tõenäosuste summana  $H_0$  kehtides. Ekstreemsemate või sama ekstreemsete juhtude kui  $X_0$  hindamiseks  $H_0$  kehtides kasutatakse Waldi teststatistikut. Waldi statistik ei ole ainuke teststatistik, millega saab jõuda nõutud tulemuseni. Kõige üldisem võimalus oleks kasutada teststatistikuna katse tulemusena realiseerunud sagedustabeli tekkimise tõenäosust ning olulisuse tõenäosuse leidmiseks summerida kõigi nende sagedustabelite tekkimiste tõenäosused, mis on võrdsed või väiksemad defineeritud teststatistiku väärtusest. Meetod on lihtsasti interpreteeritav, kuid binoomjaotuse tõenäosusvalemist tulenevalt töömahukam kui Waldi teststatistiku rakendamine. Veel oleks võimalik kasutada ka Santner'i ja Snell'i täpse testis rakendatud teststatistikut  $D$ , mis hindab sagedustabeli tekkimise ekstreemseks või veel ekstreemsemaks sõltuvalt tabeli gruppide põhjal saadud üldkogumi hinnangute vahest  $\hat{p}_v - \hat{p}_k$ . Kuigi viimase teststatistiku kasutamisel testi I liiki vea tegemise tõenäosus on väiksem Waldi teststatistiku rakendamisel saadust, on tema kasutamisel leitud testide võimsused valimi kasvades tunduvalt madalamad Waldi teststatistikuga leitudest ning kohati ka väiksemad Fisheri testi omast (D. V. Mehrotra, I. S. F. Chan ja R. L. Berger, 2003). Kuna töö üks eesmärkidest on leida suurimaid lahknevusi tingimusteta ja tingimusliku täpse testi võimsuste vahel, siis sellest tulenevalt osutub valituks Waldi teststatistik. Uuritavate sagedustabelite puhul on hinnatavaks parameetiks tõenäosuste  $p_k$  ja  $p_v$  vahe  $p_k - p_v$ , sellisel juhul lähtuvalt näiteandmestikust omandab Waldi teststatistik kuju:

$$T(X) = \frac{\hat{p}_k - \hat{p}_v}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{15} + \frac{1}{15}\right)}}, \quad (4)$$

kus  $\hat{p}_k = \frac{x_{1a}}{15}$ ,  $\hat{p}_v = \frac{x_{1b}}{15}$  ja  $\hat{p} = \frac{(x_{1a} + x_{1b})}{30}$  ning  $x_{1a} \in \{1, 2, \dots, 15\}$  ja  $x_{1b} \in \{1, 2, \dots, 15\}$ .

Äärmuslikel juhtudel, kus kas  $x_{1a}$  ja  $x_{1b}$  omavad mõlemad samaaegselt väärtust 0 või  $x_{1a} + x_{1b} = N$ , pole võimalik Waldi teststatistikut leida, kuna tekib nulliga jagamine. Testi seisukohalt tähendaks see sama jaotust mõlemas grupis, ehk tulemus ei sõltuks sellest, kas antakse päris ravimit või platseebot – haigestunud on mõlemas grupis ikka sama palju, mis omakorda viitab ravimi mitte toimimisele ning kindlasti ei saaks vastu võtta alternatiivset hüpoteesi. Sellistes situatsioonides võib Waldi teststatistiku võtta võrdseks nulliga, et tema mõju ekstreemsete väärtuste leidmisel oleks minimaalne.

Mistahes üldkogumi parameetri  $p$  väärtuse korral on eelpool kirjeldatud ühepoolse testi täpne  $p$ -väärtus arvutatav valemist:

$$p_{\text{väärtus}}(p) = \sum_{T(X) \geq T(X_0)} P(X|p). \quad (5)$$

Kahepoolse  $p$ -väärtuse leidmiseks on valem sarnane. Ainuke erinevus seisneb selles, et summa tuleb leida tõenäosustest, mille korral statistik  $T(X)$  on oma absoluutväärtuse poolest sama suur või veel suurem kui statistik  $T(X_0)$ . Barnardi testi kirjeldus jätkub ühepoolset  $p$ -väärtust silmas pidades.

Hoolimata valemi (5) lihtsasti mõisteteavast kujust, eeldab ta suuruse  $p$  teadmist, kuid praktikas enamasti pole võimalik nõutud parameetri täpset väärtust välja arvutada. Seega tuleks teda hinnata. Tänu parameetri  $p$  hindamisele saab ka otsitav  $p$ -väärtus hinnangulise väärtuse ning vaadeldavat testi ei saaks lugeda enam täpseks testiks.

Tabel 2. Vaktsiini mõju patsientidele (näiteandmestik)

| Haigestunud | Manustatud ravim |           | Kokku |
|-------------|------------------|-----------|-------|
|             | Vaktsiin         | Platseebo |       |
| Jah         | 7 (47%)          | 12 (80%)  | 19    |
| Ei          | 8 (53%)          | 3 (20%)   | 11    |
| Kokku       | 15               | 15        | 30    |



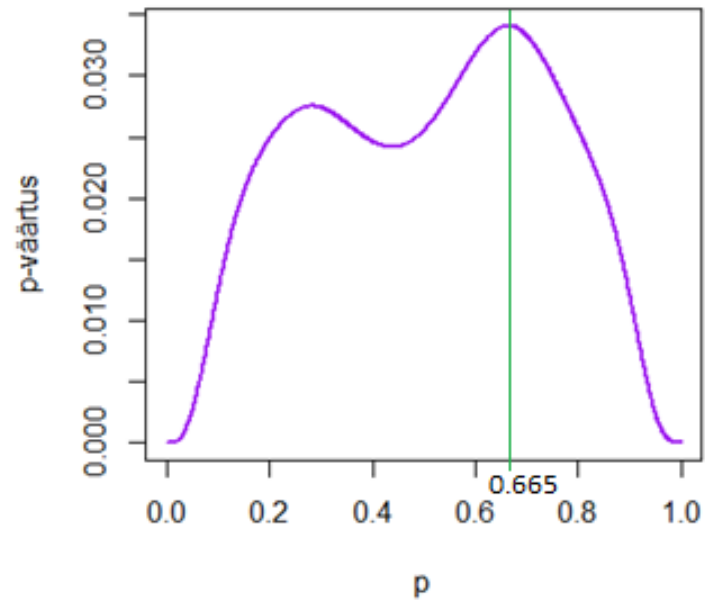
G. A. Barnard pakkus välja mitte hinnata parameetrit  $p$ , vaid leida selline parameetri  $p$  väärtus, mille korral valemist (5) tulenev olulisuse tõenäosus on kõige suurem, ehk leida suurim võimalik  $p$ -väärtus maksimeerimise teel üle parameetri  $p$  määramispiirkonna nullist üheni. Seega sõltumata sellest, milline iganes peaks üldkogumi tegelik  $p$  väärtus olema, ei saa  $p$ -väärtus olla suurem eelpool kirjeldatud suurimast  $p$ -väärtusest. Ehk kui võtta vastu alternatiivne hüpotees lähtuvalt suurimast olulisuse tõenäosusest, siis tegelik  $p$  väärtus garanteerib sama või veelgi väiksema I liiki vea tegemise tõenäosuse. Barnardi täpne  $p$ -väärtus defineerituna valemi kujul:

$$P_B := \sup\{p_{\text{väärtus}}(p) : p \in (0,1)\}. \quad (6)$$

Eelnevalt toodud näiteandmestiku puhul tuleb Waldi teststatistiku  $T(X_0)$  väärtuseks 1,894. Seega ühepoolse täpse  $p$ -väärtuse leidmiseks tuleb summeerida kõik selliste tabelite  $X$  tekkimise tõenäosused, mille korral tekkinud tabeli põhjal leitud Waldi statistik  $T(X)$  on kas suurem või sama suur kui  $T(X_0)$ . Juhul, kui toodud näites oleks täpne  $p$  väärtus teada, siis  $p$ -väärtuse saaks leida valemiga:

$$p_{\text{väärtus}}(p) = \sum_{T(X) \geq 1,894} \binom{15}{x_{1a}} \binom{15}{x_{1b}} p^{x_a+x_b} p^{(30-x_{1a}-x_{1b})}. \quad (7)$$

Kuna täpne  $p$  väärtus pole teda, ning tema hindamine ei annaks soovitud täpset  $p$ -väärtus, siis tuleb ta leida maksimeerides võrdust (7) üle  $p$  määramispiirkonna. Numbriliste meetodite abil maksimeerides väärtust  $p$  lõigul nullist üheni saavutab võrdus (7) suurima  $p$ -väärtuse 0,034  $p$  väärtuse 0,665 juures (vaata joonis 1). Järelikult olenemata üldkogumi  $p$  tegelikust väärtusest vaadeldavas näiteandmestikus, ei saa testi ühepoolne täpne  $p$ -väärtus ületada leitud suurimat  $p$ -väärtust 0,034. Seega tõenäosus teha I liiki viga peab olema 0,034 või väiksem.



Joonis 1.  $p$ -väärtus sõltuvalt tõenäosusest  $p$  lõigus  $[0,1]$  (suurim  $p$ -väärtus on märgitud sirgega)

## Fisheri täpne test

1925. aastal avaldatud Fisheri täpne test sai nime tema looja Ronald Aylmer Fisheri järgi (Wikipedia, 2013c). Siiani on Fisheri test olnud üks levinumaid teste, mida praktikas kasutatakse. Testi populaarsus on suur, kuna ta on esindatud enamikes statistika tarkvarades ning nõuab väheseid eeldusi. Hoolimata tõsiasjast, et Fisheri täpne test ei esita nõudeid valimi suurusele, kasutatakse vaadeldavat testi peamiselt väikeste valimite testimiseks. Testi tähelepanuväärseks eeliseks on tingimuslikkus, mis elimineerib parameetri hindamisest tuleneva ebatäpsuse. Teisalt tagab antud omadus testile liigse konservatiivsuse, ehk testile on omane jääda pigem nullhüpoteesi juurde.

## Hüpergeomeetriline jaotus

Hüpergeomeetriline jaotus  $X \sim HG(N, n, p)$  on diskreetne jaotus, mille tõenäosusfunktsioon

$$P(X = k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}} \quad (8)$$

kirjeldab tõenäosust  $k$  sündmuse toimumiseks suurusega  $n$  ilma tagasipanekuta valimis ehk iga järgneva sündmuse toimumine leiab aset eelnevast valimist ühe võrra väiksemas objektide hulgas. Valim  $n$  on omakorda saadud üldkogumist suurusega  $N$ ,  $n \leq N$ . Hüpergeomeetrilise jaotuse erijuhuks on  $n = 1$ , mille korral vaadeldav jaotus on võrdväärne Bernoulli jaotusega. (Traat. I, 2006, lk. 60)

## Fisheri täpse testi tööpõhimõte

Käesolev alampeatükk põhineb Mehta ja Senchaudhuri (2003) artiklil. Fisheri testi teoreetilise sisu kirjeldamiseks kasutatakse samu  $2 \times 2$  mõõtmelisi sagedustabeleid ja näitekatset, mis on kirjeldatud eelmises peatükis (vaata tabel 1 ja tabel 2). Barnardi testi tööpõhimõtte kirjeldusest tulenevalt kasutakse võrdsete tingimuste säilitamiseks ka Fisheri testi puhul teststatistikuna Waldi statistikut.

Fisheri testis vabanetakse parameetri  $p$  hindamisest tingimusel, et  $p$ -väärtuse leidmisel lähtutakse ainult katsest saadud tulemustest, ehk vaadatakse ainult neid tabelite kombinatsioone, kus sündmuste toimumiste hulk on alati sama. Teisisõnu Fisheri test kasutab  $p$ -väärtuse arvutamiseks ainult neid sagedustabeleid, kus haigestunute hulk platseebot saanute

ja vaktsiini saanute seas kokku on võrdne vaadeldud andmestikus esineva haigete hulga. Näiteandmestiku puhul on haigestunuid kokku  $x_{11} + x_{12} = 19$  (vaata tabel 2) ning testis kasutatavate tabelite hulk:

$$\Gamma(19) = \{X: X \in \Gamma \text{ ja } x_{11} + x_{12} = 19\}. \quad (9)$$

Fisher väitis, et nullhüpoteesi kehtides on samade haigete hulga tabelite juhuslikud kombinatsioonid hüpergeomeerilise jaotusega. Antud olukorras võib kõigi võimalike tabelite hulka vaadata kui üldkogumit, millest on võetud tingimuse alusel tabelite valim, milles uuritakse kindlate sündmuste arvuga tabelite saamise tõenäosusi. Seega eelnevat kirjeldatud kitsendustega juhulike tabelite esinemiste tõenäosused avalduvad üldisest valemist:

$$P(X|X \in \Gamma(s)) = \frac{\prod_{i=1}^m \binom{n_i}{x_{1i}}}{\binom{N}{s}}, \quad (10)$$

kus  $m$  on katses osalenud gruppide arv,  $s$  on saadud katses sündmuste toimumiste koguarv ehk  $s = x_{11} + x_{12} + \dots + x_{1m}$ . Kirjeldatud valemist on näha, et tabeli saamise tõenäosuse arvutamiseks pole vaja kasutada parameetrite hinnanguid. Täpse ühepoolse  $p$ -väärtuse leidmiseks tuleb summeerida kõik sellised eelnevast valemist (10) tulenevate tabelite tekkimiste tõenäosused nullhüpoteesi kehtides, mille põhjal arvutatud Waldi teststatistiku väärtus  $T(X)$  on sama suur ehk ekstreemne või veel suurem ehk ekstreemsem kui näiteandmestiku põhjal leitud teststatistiku väärtus. Seega  $p$ -väärtus on arvatav valemist:

$$p\text{väärtus} = \sum_{T(X) \geq T(X_0)} P(X|X \in \Gamma(s)), \quad (11)$$

kus  $T(X_0)$  on näiteandmestiku põhjal leitud Waldi teststatistiku väärtus. Nii nagu ka Barnardi testi kirjeldavas peatükis, on ka praegu näiteandmestiku põhjal leitud teststatistiku väärtus endiselt 1,894. Arvestades, et näiteandmestikus on haigestunute arv 19, saab valemist (11) rakendada välja arutada täpse ühepoolse  $p$ -väärtuse, milleks on 0,0641. Kui kasutada olulisuse nivood 0,05, siis Fisheri täpne test, erinevalt Barnardi täpset testist, jääks nullhüpoteesi juurde.

## Barnardi testi ja Fisheri testi võimsus

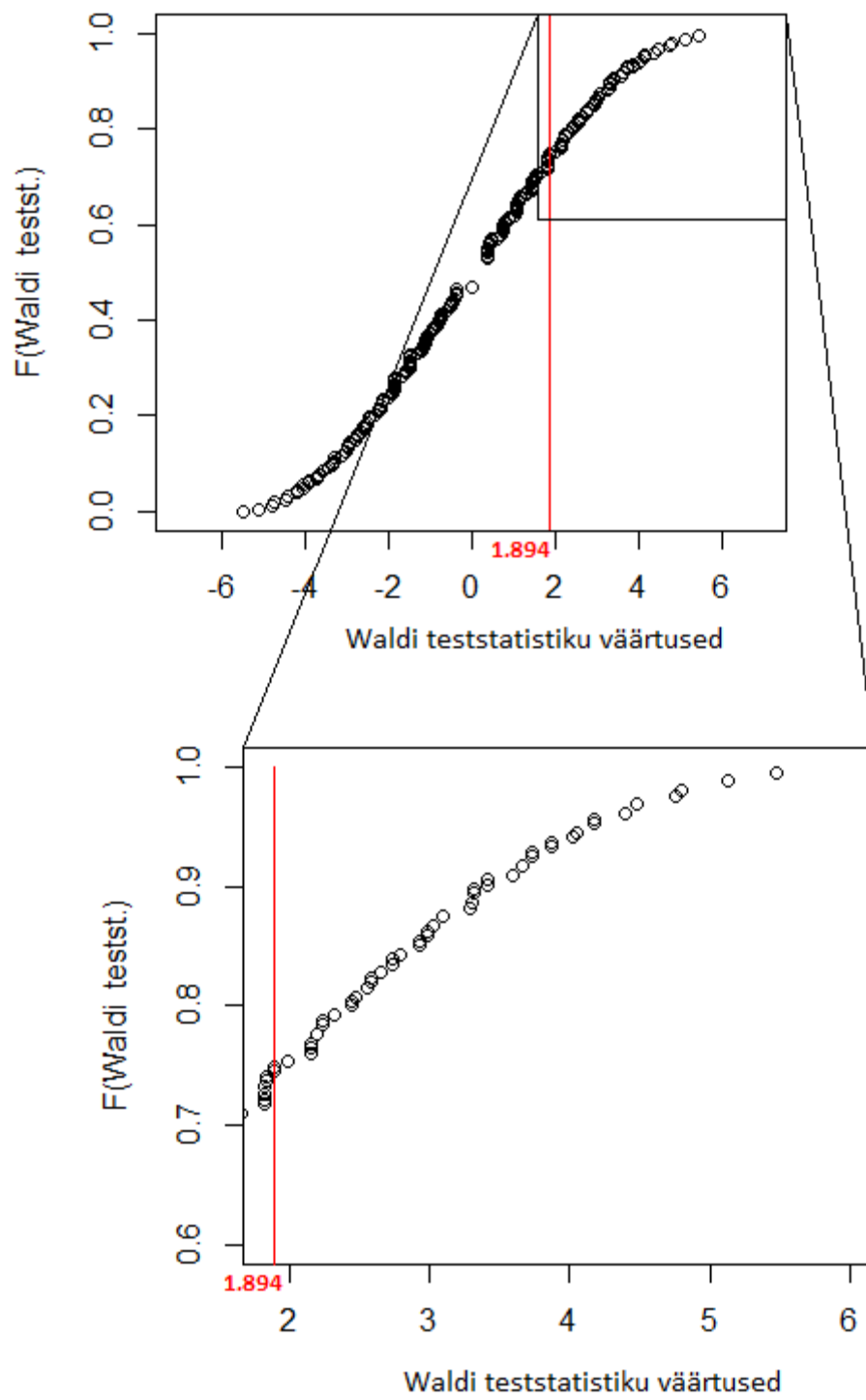
Eelnevas kahes peatükis leiti näiteandmestiku jaoks nii Barnardi kui Fisheri täpse testi ühepoolne  $p$ -väärtus, kusjuures Barnardi täpne test vastupidiselt Fisheri omale lubas usaldusnivool 0,05 vastu võtta alternatiivse hüpoteesi, ehk tõestada vaktsiini immuunsust tekitav mõju. Järgnevates alapeatükkides püütaksegi välja selgitada kumma testi tulemuse kasuks tuleks otsustada. Siinkohal on peamiseks otsustamise kriteeriumiteks karakteristikud: statistilise testi konservatiivsus ja võimsus, millest viimane näitab tõenäosust antud katseplaani korral tõestada alternatiivne hüpotees.

### Testi diskreetsus

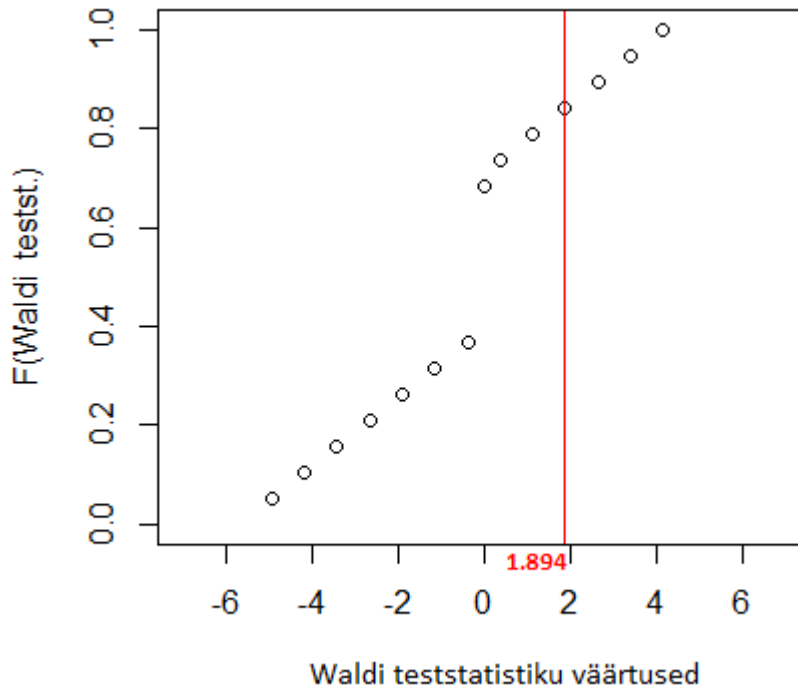
Barnardi ja Fisheri täpse testi puhul  $p$ -väärtuse leidmiseks liidetakse kokku kõigi selliste sagedustabelite tekkimiste tõenäosused nullhüpoteesid kehtides, mille korral  $T(X)$  on kas suurem või sama suur kui  $T(X_0)$ . Liites kokku kõik kirjeldatud tingimuse alusel selekteeritud tabelid, selgub, et  $p$ -väärtus ei saavuta mistahes etteantud andmestiku korral pidevat jaotust väärtuse 0,05 ümbruses või mistahes teise olulisusenivoo poolt määratud väärtuse ümbruses. Seda põhjusel, et testides kasutatud Waldi teststatistik  $T(X)$  on diskreetse jaotusega ehk  $p$ -väärtus arvutatakse lõpliku arvu tabelite tekkimiste tõenäosuste põhjal. Kui Waldi teststatistikute hulk on väike, on statistikud oma määramispiirkonnas hõredalt jaotunud ning ei pruugi sattuda nõutud olulisusenivood kirjeldava teststatistiku väärtuse lähedale, ehk ei uurita  $p$ -väärtusi olulisusenivoo lähedal. Sellel põhjusel on nii Barnardi testi, kui ka Fisheri testi poolt leitud  $p$ -väärtused konservatiivsemad ehk mõnevõrra kõrgemad kui mõne pidevat teststatistikut kasutava testi korral, sest andmestiku põhjal leitud kriitiline Waldi statistik  $T(X_0)$  on olulisusenivoost tuleneva otsuse määramiseks kas liiga suur või liiga väike (sõltuvalt uuritava probleemi püstitusest). Järelikult mida suurem on kriitilises piirkonnas asuvate Waldi statistikute hulk  $\tau_k$ , seda täpsemalt on võimalik nende põhjal  $p$ -väärtust arvutada.

Fisheri täpne test kasutab  $p$ -väärtuse leidmiseks ainult nende sagedustabelite tekkimiste tõenäosusi, mille korral on nii kontrollgrupi, kui ka eksperimentgrupi sündmuste esinemiste summa võrdne katse tulemusena saadud sagedustabeli omaga. Nõutud tingimusest tulenevalt väheneb ka diskreetsust mõjutavate teststatistikute hulk võrreldes Barnardi testiga, kus arvutatakse välja kõigi võimalike sagedustabelite teststatistikud. Fisheri testi puhul on

teststatistikute hulga diskreetsus seotud kontrollgrupi ja eksperimentgrupi sündmuste esinemiste summaga  $s$ . Nulli või valimi mahu lähedase  $s$  korral on vaadeldavate tabelite hulk väike ning sellest tulenevalt on ka väike nende põhjal leitud Waldi teststatistikute hulk. Kui töös vaadeldud näite puhul oleks  $s$  võrdne ühega, arvutataks teststatistikud ainult kahe tabeli põhjal, kas üks haigestunu vaktsiini saanute grupis ning null haigestunut platseebot saanute grupis või vastupidi. Analoogne situatsioon leiaks aset tervete hulgas juhul, kui  $s$  oleks valimi mõistes suur. Selgub, et Fisheri testi diskreetsus fikseeritud gruppide arvu ja sama valimi mahu juures väheneb ainult juhul, kui väärtus  $s$  läheneb poolele valimi mahust, sest tabelite kombinatsioon, kus sündmuste toimumiste summa on ligikaudu pool valimi mahust, on rohkem. Näiteandmestiku puhul, lähtuvalt  $s$  väärtusest 19, on  $p$ -väärtus arvatud ainult nelja sama või veel ekstreemsema tabeli tekkimise tõenäosuste summana (vaata joonis 3). Kuna arv  $s$  on määratud tänu katse tulemusele, siis üks võimalus Fisheri testi diskreetsust vähendada oleks valimimahu suurendamine. Kui esialgne katse on hea plaani alusel koostatud, siis valimimahu suurenedes jäävad osakaalud gruppides samaks ning Fisheri testi diskreetsus väheneb, kuna teststatistiku väärtuste hulk suureneb. Uurides teststatistiku jaotust Barnardi testi puhul selgub, et teststatistiku väärtuste hulk katab kõik Fisheri testi teststatistiku väärtused ning neile lisanduvad veel ka väärtused, mis on saadud kõigi võimalike juhuslike tabelite põhjal. Näiteandmestikus, kasutades Barnardi testi, arvutatakse olulisuse tõenäosus 64 tabeli põhjal (vaata joonis 2). Seega Barnardi testi diskreetsus on oluliselt väiksem Fisheri testi omast. Vaadeldav eelis tulenebki sellest, et Barnardi testi puhul ei ole seatud tingimust leida teststatistiku väärtus ainult nende tabelite põhjal, mille sündmuste õnnestumiste summa on fikseeritud. Tänu sellele puudub ka seos, et erinevad sündmuste toimumiste summad  $s$  mõjutaksid testi diskreetsust. On ilmne, et valimi suurenedes ka Barnardi testi diskreetsus väheneb, kuna erinevate võimalike sagedustabelite hulk suureneb veelgi, kuid Barnardi testi diskreetsus on valimi suurusel vähem mõjutatud kui Fisheri täpse testi diskreetsus.



Joonis 2. Waldi teststatistiku väärtuste jaotus Barnardi testi korral



Joonis 3. Waldi teststatistiku väärtuste jaotus Fisheri testi korral

### Suurematele kui 2 x 2 sagedustabelitele rakendatud testide diskreetus

Objektide arvu suurenemine grupis ei ole ainus vaadeldud testide diskreetuse vähenemise põhjus. Testi konservatiivsus väheneb ka tabeli mõõtmete kasvades. Grupisisene binoomjaotus kehtib üksnes siis, kui grupp koosneb binaarse tunnuse väärtustest. Seega ainus võimalus tabeli mõõtmete suurenemiseks on gruppide arvu kasv ehk uurimise alla saavad kuuluda ainult  $2 \times m$  mõõtmelised sagedustabelid, kus  $m \in \mathbb{N}$ . Suvalise  $2 \times m$  mõõtmelise tabeli saamise tõenäosused on arvutatavad valemi (3') ja (10) abil. Fisheri testi puhul tuleb sündmuste toimumiste summa  $s$   $2 \times m$  mõõtmelise tabeli korral leida  $m$  erineva liidetava põhjal, millest tulenevalt testis läbivaadatavate tabelite hulk kasvab, kuna sündmuste toimumiste kombinatsioone, mis üle kõigi gruppide annavad kokku summaks  $s$ , on rohkem. Fisheri testiga  $p$ -väärtuse arvutamiseks sobivate tabelite hulga kasvades, kasvab ka teststatistiku väärtuste hulk ning seeläbi väheneb ka testi diskreetus. Kuna Barnardi testis tabelite põhjal leitud teststatistiku väärtuste hulk katab Fisheri testis kasutatavate oma, siis tabeli mõõtmete suurenedes väheneb ka Barnardi testi diskreetus. Barnardi testi diskreetuse eelis Fisheri testi ees tabeli mõõtmete kasvades järjest väheneb, sest praktikas nõutud diskreetusest tulenev ebatäpsus muutub ebaoluliseks.



Väikeste valimite korral on hüpoteeside kontrollimises sobilikum kasutada Fisheri täpse testi asemel Barnardi täpset testi, sest tema diskreetus on oluliselt madalam. Sama tulemus kehtib ka olukorras, kui uuritava andmestiku põhjal saadud sagedustabeli mõõtmed on miniatuursed.

## Tõepärasuhte test

Kahe erineva testi omaduste võrdlemisel saab öelda, kas üks on parem kui teine või on nad oma omadustelt võrdsed, kuid see ei anna terviklikku pilti ühe või teise testi sobivuseks mingite hüpoteeside kontrollimiseks. Võib tekkida olukord, kus üks test on küll omadustelt parem kui teine, kuid kumbki neist ei sobi nõutud hüpoteesi kontrolliks. Siiani on võrreldud ainult kahte täpset testi. Et võrdlus ei jääks ainult nendest lähtuvaks võrreldakse neid ka tõepärasuhte testiga.

Definitsioon 1. Olgu antud valim  $x_1, x_2, \dots, x_n$  jaotuses  $F(x, \theta)$ , mis võib olla kas pidev või diskreetne. Tõepärafunktsiooniks nimetatakse avaldist:

$$L(\theta) = \begin{cases} f(x_1; \theta) * f(x_2; \theta) * \dots * f(x_n; \theta) \\ p(x_1; \theta) * p(x_2; \theta) * \dots * p(x_n; \theta) \end{cases}, \quad (12)$$

kus  $f(x; \theta)$  on jaotuse  $F$  tihedusfunktsioon ja  $p(x; \theta)$  on jaotuse  $F$  tõenäosusfunktsioon ning  $\theta$  üldkogumi parameeter. (Traat I., 2006, lk. 42)

Definitsioon 2. Väärtust  $\hat{\theta}$  parameeterruumis  $A$ , mille korral  $L(\theta)$  saavutab maksimaalse väärtuse, nimetatakse parameetri  $\theta$  suurima tõepära hinnanguks:

$$L(\hat{\theta}) = \max_{\theta \in A} L(\theta). \quad (13)$$

(Traat I., 2006, lk. 43)

Tõepärasuhte testi teststatistikuks on tõepära suhe. Teststatistiku nimetajas on valimi tõepärafunktsioon ning lugejas valimis tõepärafunktsioon nullhüpoteesi eeldusel. Kui kasutada esimeses peatükis toodud 2 x 2 sagedustabelit (vaata tabel 1) ning seal defineeritud muutujaid, siis ühistõenäosuse  $\pi_{ij}$  ehk sattuda kummagi grupi kas haigestunute või immuunsete hulka suurima tõepära hinnanguks on  $\hat{\pi}_{ij} = \frac{x_{ij}}{N}$ . Vaadeldud valimi saamise tõenäosuseks on sellisel juhul:

$$\prod_{i=1}^2 \prod_{j=1}^m \pi_{ij}^{x_{ij}}, \quad (14)$$

kus  $m$  tähistab tabeli veergude ehk gruppide arvu. Asendades valemis  $\pi_{ij}$  hinnanguga  $\hat{\pi}_{ij}$  saadakse tõepärasuhte statistiku nimetaja. Kui uuritava andmestiku kirjeldavate tunnuste põhjal saadud gruppide suurused on juhuslikud, siis nullhüpoteesi eeldusel  $\pi_{ij} = \pi_i \pi_j$ , kus  $\pi_i$  ja  $\pi_j$  tähistavad marginaaljaotuste ehk haigestumise ja mittehaigestumise ning vaktsineeritute ja mittevaktsineeritute gruppi sattumise tõenäosusi, mille suurima tõepära hinnanguks on  $\hat{\pi}_{ij} = \frac{n_{+j}n_{i+}}{N^2}$ . Asendades saadud tulemus valemis (14) oleva parameetri  $\pi_{ij}$  asemele, saadakse teststatistiku lugeja. Seega tõepärasuhte teststatistik avaldub kujul:

$$\Lambda = \frac{\prod_{i=1}^2 \prod_{j=1}^m \left(\frac{n_{+j}n_{i+}}{N^2}\right)^{x_{ij}}}{\prod_{i=1}^2 \prod_{j=1}^m \left(\frac{x_{ij}}{N}\right)^{x_{ij}}}. \quad (15)$$

Kui nullhüpotees on väär, saavutab tõepärafunktsioon maksimumi saadud statistiku nimetajas ning mida väiksem on  $\Lambda$ , seda põhjendatum on nullhüpoteesi kummutamine. Tõepärasuhte teststatistiku logaritmime ning  $-2$  korrutamise tagajärjel on saadud tulemus nullhüpoteesi kehtimisel asümptootiliselt hii-ruut jaotusega, mille vabadusastmete arvuks on nimetajas ja lugejas oleva mudeli vabadusastmete arvu vahe (Traat. I, 2013, lk. 26; Wikipedia, 2013d).

Vaadates tabeleid, kus tunnuse alusel tekkinud gruppide suurused on fikseeritud, kehtib nullhüpoteesi eeldusel kas  $\pi_{ij} = \pi_i$  või  $\pi_{ij} = \pi_j$ . Valik sõltub uuritava probleemi püstitusest. Saadud tulemuste hinnangud avalduvad vastavalt kujul  $\hat{\pi}_{ij} = \frac{n_{+j}}{N}$  ja  $\hat{\pi}_{ij} = \frac{n_{i+}}{N}$ . Tõepärasuhte teststatistiku saamiseks tuleb sõltuvalt tabeli tüübist leitud hinnangud asendada valemi (15) lugejasse. Saadud suhte logaritmitud ning  $-2$  korrutatud kuju on samuti asümptootiliselt hii-ruut jaotusega.

## Võimsusfunktsioon

Ühe ja sama hüpoteesipaari kontrollimiseks saab konstrueerida palju teste. Parim test on see, mille I liiki vea tegemise tõenäosus, ehk tõenäosus võtta vastu  $H_1$  kuigi kehtib  $H_0$ , on väike ja mis suure tõenäosusega kummutab  $H_0$ , kui  $H_1$  on õige. Parema testi valimiseks võib kasutada testide võimsusfunktsioonide võrdlust. Testi võimsusfunktsioon  $h(\theta)$  on testi kummutamise tõenäosus tingimusel, et väärtus  $\theta$  on parameetri õige väärtus üldkogumis:

$$h(\theta) = P(\text{kummtada } H_0 | \theta) \quad (16)$$

ehk alternatiivselt

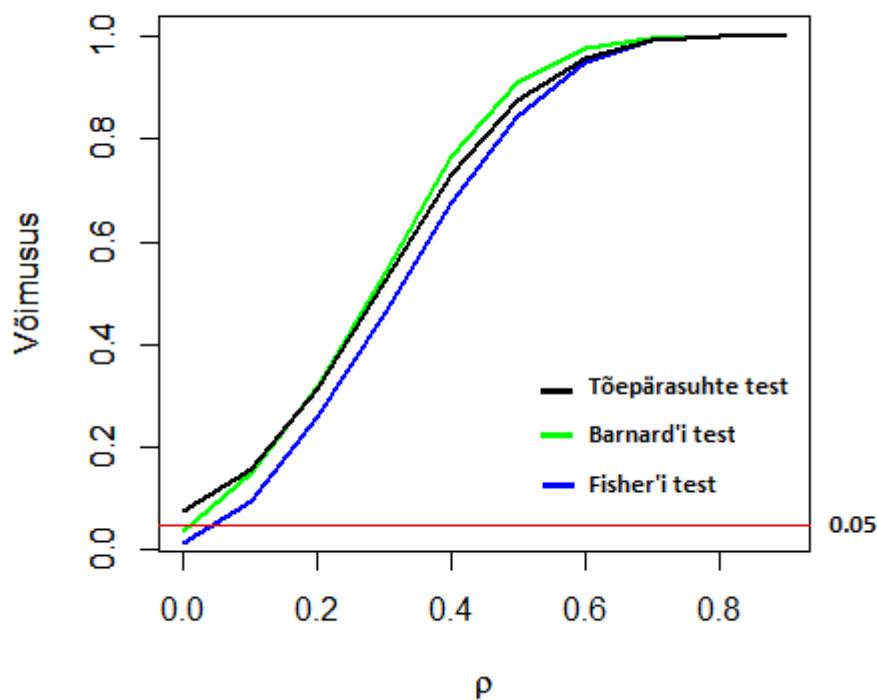
$$h(\theta) = P(T(X) \in C | \theta), \quad (17)$$

kus  $C$  on teststatistiku kriitiline piirkond. Seega hea testi korral on  $h(\theta)$  väike, kui  $\theta \in H_0$  ning suur, kui  $\theta \in H_1$ . Kuna  $h(\theta)$  puhul on tegemist tõenäosusega, siis tema määramispiirkond jääb vahemikku nullist üheni. (Traat I., 2006, lk. 113)

Nii Barnardi, Fisheri kui ka tõepärasuhte testi võimsuste leidmiseks uuritavate 2 x 2 sagedustabelite jaoks kasutatakse võimsusfunktsiooni parameetrina kontrollgrupi sündmuste toimumiste tõenäosuste ning eksperimentgrupi sündmuste toimumiste tõenäosuste vahet  $\rho = p_k - p_v$ . Kuna töö eesmärk on uurida testide omadusi väikeste valimite korral, siis sellest tulenevalt kasutatakse testide võrdlemiseks valimi suurust, mis on võrdne esimeses peatükis kirjeldatud andmestiku omaga (vaata tabel 2). Kui kontrollgrupis ja eksperimentgrupis on sündmuste toimumiste tõenäosused võrdsed on  $\rho$  võrdne nulliga, mis viitab nullhüpooteesi kehtimisele ning võimsusfunktsiooni väärtused peaksid tulema väikesed. Üldkogumi tõenäosuste erinevuse kasvades suureneb ka parameetri  $\rho$  absoluutväärtus ning parema testi puhul peaksid võimsusfunktsiooni väärtused olema suuremad. Järgnevalt uuritakse testide võimsusi kolme eritüüpi 2 x 2 sagedustabelitel: juhuslike marginaalidega, ühe tunnuse korral fikseeritud marginaalidega ning kõigi tunnuste puhul fikseeritud marginaalidega.

Esimesel juhul genereeritakse iga parameetri  $\rho \in \{0; 0,1; \dots; 1\}$  väärtuse kohta 100000 juhuslikku sagedustabelit, kus nõutud tunnust omavate objektide arv grupis on võrdne eelnevalt toodud andmestiku omaga, kus nii vaktsiini saanute kui ka mittesaanute grupis oli 15 patsienti. Võimsusfunktsiooni väärtused leitakse saadud tabelitele testi rakendamisel testi poolt alternatiivse hüpooteesi kasuks otsustatud tabelite hulga ja vaadeldud 100000 tabeli jagatisena. On näha, et Barnardi testi puhul võimsusfunktsioonide väärtused on keskmiselt 0,007 kõrgemad tõepärasuhte testi omadest ning keskmiselt 0,04 võrra kõrgemad Fisheri testi omadest (vaata joonis 4). Lisaks on tulemustest märgata ka eelnevas alampeatükis kirjeldatud testide diskreetsuste erinevused. Fisheri testi ehk kolmest vaadeldud testist konservatiivseima puhul langevad võimsusfunktsiooni väärtused kõige madalamale. Täpsete testide diskreetsuse omadusest tulenevalt langevad nende testide võimsusfunktsioonide väärtused

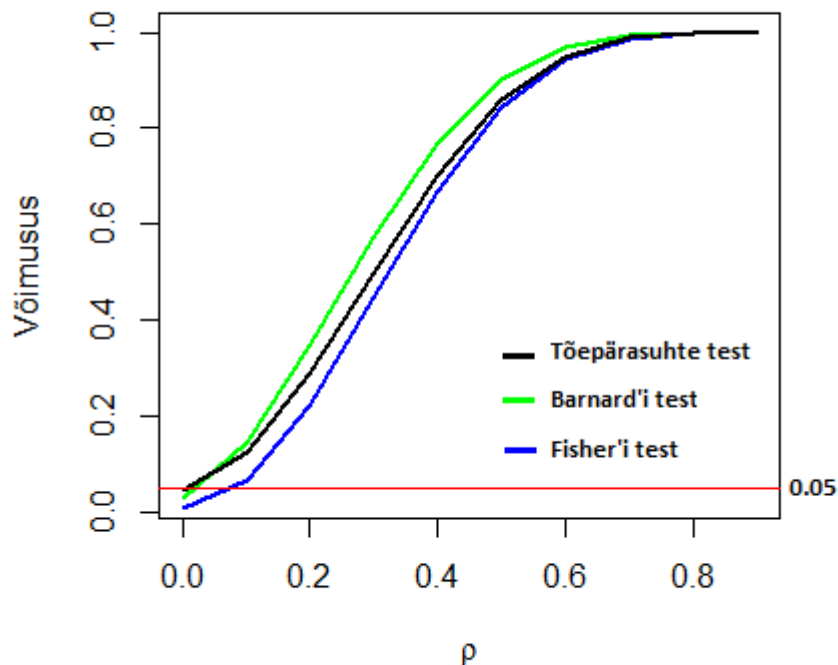
kohati ka alla 0,05. Oluliselt ohtlikumaks tulemuseks siinkohal on tõepärasuhte testi võimsusfunktsioon, mille väärtused kordagi ei lange väärtuseni 0,05 ega madalamale. Järelikult leidub andmestikke, kus olulisuse nivool 0,05 tuleks jääda nullhüpooteesi juurde, kuid tõepärasuhte test võtab vastu alternatiivse hüpooteesi ehk tema poolt raporteeritav p-väärtus tuleb tegelikust madalam. Tekkinud olukord on põhjustatud väikesest valimi mahust. Seega tõepärasuhte testi ei tohiks kasutada vaadeldavate andmestike põhjal hüpooteeside kontrollimiseks, mis seda enam rõhutab vajadust väikese valimi korral korrektselt töötavate testide järele.



Joonis 4. Testide võimsused ühe fikseeritud marginaali korral; 15 objekti grupis

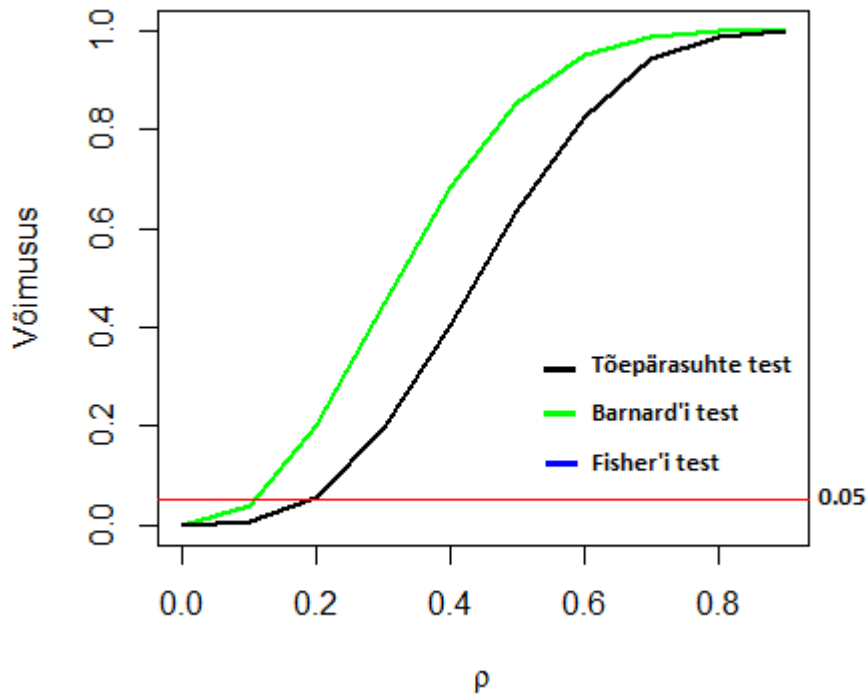
Teisel juhul genereeritakse iga parameetri  $\rho \in \{0; 0,1; \dots; 1\}$  väärtuse kohta 100000 juhuslikku sagedustabelit, kus  $2 \times 2$  sagedustabeli marginaalid on juhuslikud arvud binoomjaotusest keskväärtusega 15, et saadud tabelite valimi maht oleks sarnane näiteandmestikuga. Nii nagu ka fikseeritud grupi suuruste korral on ka juhuslike marginaalide korral Barnardi testi võimsusfunktsioonide väärtused kõrgemad kahe ülejäänud testi omadest. Keskmiselt 0,027 võrra kõrgemad tõepärasuhte testi omadest ning Fisher'i testi omadest

keskmiselt 0,054 võrra kõrgemad (vaata joonis 5). Taaskord on jooniselt näha, et täpsed testid kalduvad jääma nullhüpoteesi juurde, kuna diskreetsusest tulenevalt on nende p-väärtused sama tabeli juures kõrgemad kui pideva jaotustega teststatistikut kasutava testi korral.



Joonis 5. Testide võimsused juhulike marginaalide korral; keskmiselt 15 objekti grupis

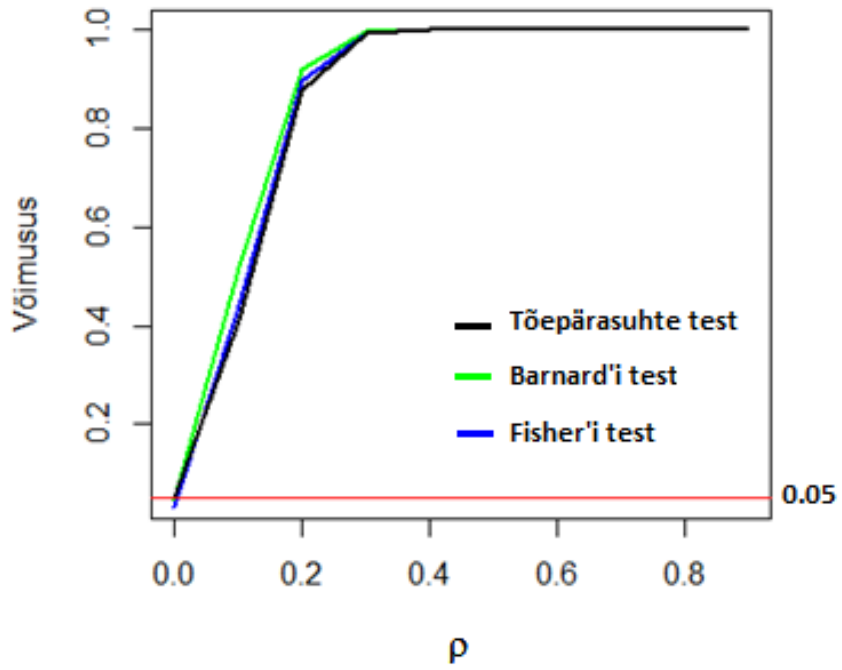
Viimasena vaadeldakse ka kolmandat tüüpi sagedustabeleid, mille marginaalid on mõlema tunnuse jaoks fikseeritud. Eelnevalt kasutatud näiteandmestiku korral pole võimalik sellist situatsiooni kirjeldada, sest haigestunute hulk peale vaktsiini või platseebo saamist on juhuslik. Terviklikuma ülevaate saamise eesmärgil uuritakse testide võimsuste käitumisi ka nende tabelite korral. Seega genereeritakse iga parameetri  $\rho \in \{0; 0,1; \dots; 1\}$  väärtuse kohta 100000 juhuslikku  $2 \times 2$  sagedustabelit, kus tabeli marginaalid mõlema tunnuse korral on fikseeritud arvuga 15. Fisheri testi ja tõepärasuhte testi tulemused langevad kokku ning Barnardi testi võimsusfunktsiooni väärtused on kummagi testi võimsustest keskimelt 0,11 võrra kõrgemad (vaata joonis 6).



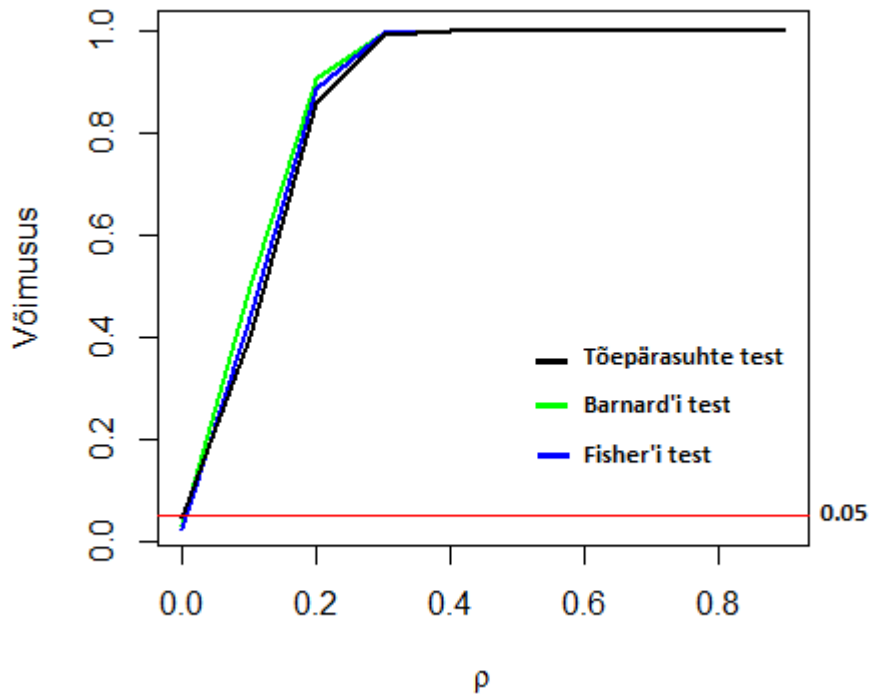
Joonis 6. Testide võimsused fikseeritud marginaalide korral; 15 objekti grupis (tõepärasuhte testi ning Fisheri testi võimsusfunktsiooni joon langevat kokku)

Korrates samu võimsusfunktsioonide leidmise meetodeid kolmel eelnevalt defineeritud 2 x 2 sagedustabelitel, kuid kus nüüd kummaski grupis on 75 või keskmiselt 75 objekti ehk valimi suuruseks on keskmiselt 150 objekti, on näha, et testide võimsusfunktsioonide erinevused on tunduvalt väiksemad kui 30 objektiga valimi korral (vt joonis 7,8,9). 1985. aastal jõudsid Suissa ja Shuster tulemusele, et tingimusteta Barnardi test omab võimsuse eelist kirjeldatud tingimuslikkusele tugineva testi ees 2 x 2 sagedustabelite juures kuni valimi mahu kasvuni 150 objekti kummaski grupis (Mehta C. R. ja Hilton J.F., 1993, lk. 91).

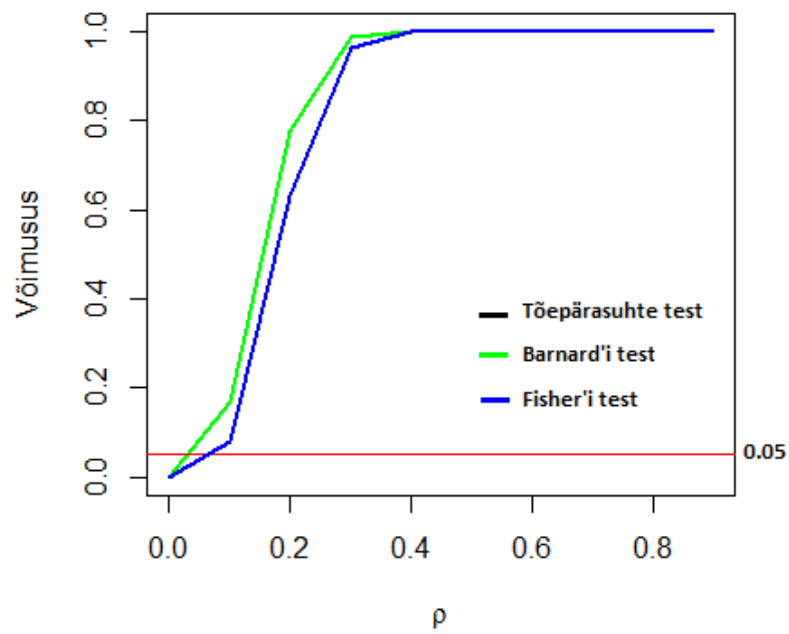
Võttes testi valiku suhtes otsustavaks karakteristikuks testi võimusus, tuleks väikeste valimitega 2 x 2 sagedustabelite põhjal hüpoteeside kontrollimiseks kasutada Barnardi täpset test.



Joonis 7. Testide võimsused ühe fikseeritud marginaali korral; 75 objekti grupis



Joonis 8. Testide võimsused juhuslike marginaalide korral; keskmiselt 75 objekti grupis



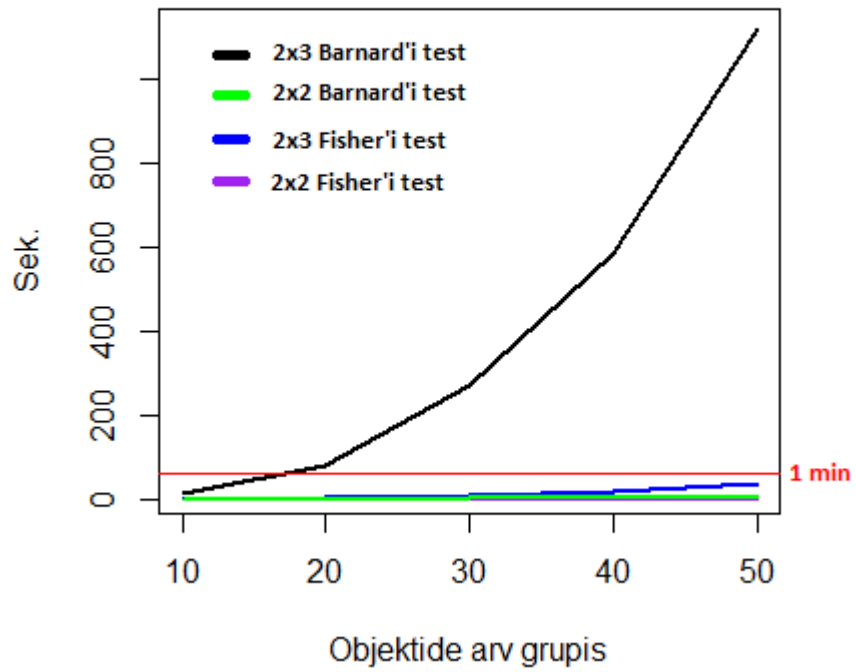
Joonis 9. Testide võimsused fikseeritud marginaalide korral; 75 objekti grupis (tõepärasuhte testi ning Fisher'i testi võimsusfunktsiooni joon langevat kokku)



## Probleemid Barnardi p-väärtuse leidmisel

Eelnevas peatükis selgus, et Barnardi testi võimsuse eelis võrreldes Fisheri testiga tabeli mõõtmete kasvades või valimi mahu suurenedes väheneb. Lisaks võimsuse eelise kadumisele ilmneb probleeme ka Barnardi testi läbiviimisega vajaliku parameetri  $p$  hindamisel. Sagedustabeli suuruse kasvades kasvab läbivaadatavate võimalike tabelite hulk eksponentsiaalselt (Mehta C. R. ja Hilton J.F., 1993, lk. 98). Näiteks kui vaatluse all on  $2 \times 2$  mõõtmeline sagedustabel, tuleb valemit (7) maksimeerida üle kõikvõimalike väärtuste  $x_{1a}$  ja  $x_{1b}$ , kuid  $2 \times 3$  mõõtmeliste sagedustabelite puhul üle kõikvõimalike väärtuste  $x_{1a}$ ,  $x_{1b}$  ja  $x_{1c}$ , kus  $x_{1c}$  on kolmandas grupis sündmuste esinemiste arv. Tabeli mõõtmete kasvades suureneb parameetrite arv veelgi ning maksimeerimise protsess muutub veelgi aega nõudvamaks. Nii Fisheri kui ka Barnardi testi korral leitakse  $2 \times 2$  juhuslike sagedustabelite tõenäosused vastavalt eelpool defineeritud valemitele. Suuremate tabelite korral ei leita kriitilisi väärtusi Waldi teststatistikut kasutades, vaid liidetakse kokku kõik sellised tõenäosused, mille korral juhusliku tabeli tekkimise tõenäosus on väiksem katse tulemusena saadud tabeli omast. Kasutades rakendustarkvaras R defineeritud koodi (vaata lisa 1) nii  $2 \times 2$  sagedustabeli kui ka  $2 \times 3$  sagedustabeli korral, kus igas grupis on 15 objekti, kulub Barnardi testi täpse ühepoolse p-väärtuse leidmiseks  $2 \times 3$  tabeli puhul 8 sekundit rohkem aega kui  $2 \times 2$  sagedustabeli korral. Joonisel 10 on toodud võrdlus  $2 \times 2$  ja  $2 \times 3$  tabeli p-väärtuse leidmiseks sõltuvalt objektide arvust grupis. Kusjuures igas grupis on objektide arv võrdne. Lisaks on joonisel näidatud ka Fisheri täpse testiga p-väärtuse leidmiseks kuluv aeg, kasutades lisa 1 toodud rakendustarkvara R-i koodi, sõltuvalt eelpool mainitud Barnardi testile seatud tingimustele. On selge, et Fisheri testi teostava algoritmi abil p-väärtuse välja arvutamine ei sõltu mitte nii palju valimi mahust kui sündmuste toimumiste kogusummast  $s$ . Kui  $s$  väärtus on võrdne poolega valimi mahust, tuleb läbi vaadata maksimaalne arv tabeleid ning välja arvutada nende tekkimise tõenäosused. Kuna selline situatsioon on kõige enam aeganõudev, siis leitakse testi läbiviimiseks kuluvad ajad just nende tabelite põhjal.

Testide tegemiseks kasutati sülearvutit protsessoriga 1,6GHz. Teste läbiviivale programmile R jagati pool protsessori töö võimsusest ehk programmi töötamise ajal tehti 800 000 arvutussammu sekundis.



Joonis 10. Testi tegemiseks kuluv aeg sõltuvalt valimi mahust

## Kokkuvõte

Käesoleva töö eesmärk oli võrrelda kahest erinevat täpset testi – Barnardi tingimusteta testi ja Fisheri tingimuslikku testi. Peamiselt kasutatakse uuritavaid teste hüpoteeside kontrollimiseks väikeste valimi mahtudega binaarsete tunnustega sagedustabelitel. Töö esimeses ja teises peatükis kirjeldati mõlema testi töö põhimõtet. Kirjelduse paremaks mõistmiseks oli peatükkidesse kaasatud 2 x 2 sagedustabelisse paigutatud näidisandmestik valimi suurusega 30 objekti. Andmestiku põhjal püstitatud hüpoteeside testimisel saavutas Barnardi täpne test madalama p-väärtuse kui Fisheri täpne test, kusjuures võttes vastu alternatiivse hüpoteesi erinevalt Fisheri testist (olulisuse nivool 0,05).

Järgnevas peatükis anti ülevaade testide diskreetsusest ning võrreldi testide võimsusi. Et võrdlus ei oleks jäänud ainult nimeteatud kahest testist lähtuvaks, võrreldi neid ka tõepärasuhte testiga. Eelnevalt toodud näiteandmestiku korral on Fisheri test oluliselt diskreetsam Barnardi testist, kuna tingimuslikkusest tulenevalt on testis kasutatavate teststatistiku väärtuste hulk kordades väiksem kui Barnardi omas. Testide võimsusi uuriti kolme eritüüpi 2 x 2 sagedustabelitel: ühe tunnuse korral fikseeritud marginaalidega, mõlema tunnuse korral juhuslike marginaalidega ning mõlema tunnuse korral fikseeritud marginaalidega. Esialgu leiti testide võimsused tabelitel valimite suurustega 15 objekti kummaski grupis ning hiljem 75 objekti kummaski grupis. 2 x 2 sagedustabelite korral kuni valimi mahu kasvuni 150 objekti kummaski grupis omab Barnardi test võimsuse eelist tingimuslikkusele tugineva Fisheri testi ees.

Töö viimases osas uuriti ühte peamist Barnardi testi läbiviimisel ilmnevat probleemi – testi tegemiseks kuluv aeg. Sagedustabeli mõõtmete suurenedes kasvab läbivaadatavate võimalike tabelite hulk eksponentsiaalselt, seega pikeneb ka testi läbiviimiseks kuluv aeg. Peatükis toodud jooniselt on näha, et 2 x 2 sagedustabelite korral kuni valimi mahuni 50 objekti kummaski grupis, kulub nii Barnardi kui Fisheri testi tegemiseks aega alla 1 minuti. Teisalt 2 x 3 sagedustabelite korral, kus igas grupis on 50 objekti, nõuab Barnardi testiga p-väärtuse leidmine aeg üle 15 minuti, samal ajal kui Fisheri testiga kulub selleks endiselt alla 1 minuti.

Viimasest kahest peatükist lähtuvalt leiab töö autor, et väikeste valimitega 2 x 2 sagedustabelite põhjal püstitatud hüpoteeside testimiseks tuleks kasutada Barnardi testi.

# Barnard's and Fisher's exact test comparison

Bachelor thesis

Edwart Ždanoviš

## Summary

The purpose of this research was to give an overview of two exact tests - Barnard's unconditional test and Fisher's conditional test, and find their pros and cons while using them to test independence between two (usually binary) variables. First two sections described both tests by using Wald's test statistics. In the thesis, an example with a sample size of 30 is presented where Barnard's exact test achieves lower p-value than Fisher's exact test and is able to reject  $H_0$  (unlike Fisher's exact test).

The next chapter discusses the discreteness problems of the test and examines the power of both tests. To get a better picture of the power comparison, the likelihood-ratio test was also included. Powers of tests were compared in three types of 2 x 2 contingency tables: fixed marginal on one variable, random marginal on both variables and fixed margins on both variables, with a sample size of 30 and later with a size of 150. Contingency tables with a sample size 30 Barnard's exact test got fairly higher powers than Fisher's exact test. Powers of the likelihood-ratio test and Barnard's exact test were similar. If sample size gained value of 150, all test's powers were similar. With a sample size as large as 150 in both groups, Barnard's exact test has greater power than Fisher's one.

The last section gives a brief overview of one of the major difficulties of Barnard's exact test – complexity of calculating test results. The summation of equation in Barnard's test would grow exponentially, if the size of the table or sample size grew. Graph shows that with 2 x 2 contingency tables up to a sample size of 50 per group, both tests are able to give p-value with less than 1 minute. However, with 2 x 3 contingency tables with a sample size of 50 per group, Barnard's test needs more than 15 minutes to give p-value, unlike Fisher's exact test which required less than 1 minute.

Based on the last two sections, the author of this research paper has decided to use Barnard's exact test for analysing the hypotheses on 2 x 2 contingency tables with small sample size.

## Kirjanduse loetelu

Mehta C. R. ja Senchaudhuri P., 2003. Conditional versus Unconditional Exact Test for Comparing Two Binomials [pdf], Saadaval: [www.cytel.com/Papers/twobinomials.pdf](http://www.cytel.com/Papers/twobinomials.pdf) [Alla laetud 27 märts 2013]

Mehrotra D. V., Chan I. S. F. ja Berger R. L., 2003. A Cautionary Note on Exact Unconditional Inference for a Difference between Two Independent Binomial Proportion, *Biometrics*, [e-ajakiri] 59(411-450), Saadaval läbi: Tartu Ülikooli Raamatukogu [Alla laetud 27 märts 2013]

Mehta C. R. ja Hilton J.F., 1993. Exact Power of Conditional and Unconditional Tests: Going Beyond 2 x 2 Contingency Table, *The American Statistician*, [e-ajakiri] 47(2), Saadaval läbi: Tartu Ülikooli Raamatukogu, Saadaval: <http://atoz.ebsco.com/Titles/1249?lang=en&lang.menu=en&lang.subject=en> [Alla laetud 27 märts 2013]

Traat, I, 2006. *Matemaatilise statistika põhikursus*. Tartu: Tartu Ülikooli Kirjastus

Traat, I, 2013. Hüpooteesid sõltumatuse kohta. Aine kvalitatiivsete andmete analüüs loengu konspekt. [pdf] (25), Saadaval läbi: Tartu Ülikooli õppeinfosüsteem, [Alla laetud 3 aprill 2013]

Wikipedia, 2013a. *Barnard's test*. [internet]. Saadaval: [http://en.wikipedia.org/wiki/Barnard%27s\\_test](http://en.wikipedia.org/wiki/Barnard%27s_test). [Alla laetud 27 märts 2013]

Wikipedia, 2013b. *Wald test*. [internet]. Saadaval: [http://en.wikipedia.org/wiki/Wald\\_test](http://en.wikipedia.org/wiki/Wald_test). [Alla laetud 27 märts 2013]

Wikipedia, 2013c. *Fisher's exact test*. [internet]. Saadaval: [http://en.wikipedia.org/wiki/Fisher%27s\\_exact\\_test](http://en.wikipedia.org/wiki/Fisher%27s_exact_test). [Alla laetud 27 märts 2013]

Wikipedia, 2013d. *Likelihood-ratio test*. [internet]. Saadaval: [http://en.wikipedia.org/wiki/Likelihood\\_ratio\\_test](http://en.wikipedia.org/wiki/Likelihood_ratio_test). [Alla laetud 15 aprill 2013]

# LISA 1

Töös kasutatud tarkvara R programmi kood:

```
require(combinat)
library(Barnard)

# Wald statistik
Wald = function(xa,xb,n1,n2){
  pi1 = xa / n1
  pi2 = xb / n2
  pik = (xa + xb) / (n1 + n2)
  if (pik*(1-pik) > 0){
    w = (pi2 - pi1) / ((pik*(1 - pik)*(1/n1
+ 1/n2))**0.5)
  } else {
    w = 0
  }
  return (w)
}

# Binoom jaotuse toenaosus valem
binoom = function(xa,xb,n1,n2,pi){
  b = choose(n1,xa) * choose(n2,xb) *
pi^(xa+xb) * (1-pi)^(n1 + n2 - xa - xb)
  return (b)
}

# Leiab kõik kombinatsioonid, kus T(X)>T(X0)
p = function(xa,xb,xc,xd,pi){
  n1 = xa + xc
  n2 = xb + xd
  w0 = Wald(xa,xb,n1,n2)
  wSumma = 0
  for (m in 0:n1){
    for (n in 0:n2){
      if (Wald(m,n,n1,n2) >= w0){
        wSumma = wSumma + binoom(m,n,n1,n2,pi)
      }
    }
  }
  return (wSumma)
}

# BARNARDI TEST
barnard = function(xa,xb,xc,xd,tapsus){
  #print(matrix(c(xa,xc,xb,xd),nrow=2, ncol=2))
  x = seq(0,1,by=tapsus)
  #y = p(xa,xb,xc,xd,x)
  #plot(x,y,type="l",xlab = "p",ylab="p-
väärtus",col="purple",lwd=2)
  #print("p-vaartus Barnardi yhepool testil")
  pi = max(p(xa,xb,xc,xd,x))
  #lines(c(pi,pi),c(0,1),col="blue")
  if (pi == "NaN"){
    pip = 1
  }
  else {
    pip = pi
    #b_voimsus(xa+xc,xb+xd,pip)
  }
  return(pip)
}
```

```
}

=====

#TOEPARASUHTE TEST (Saadud:
http://www.pmc.ucsc.edu/~mclapham/Rtips/G%20test.tx
t)

=====

#TESTIDE VOIMSUSED

voimsuseleidja1 = function(z,a,n){

  sv = seq(0,1,1,0.1)
  bvoimsus = rep(0,length(sv))
  fvoimsus = rep(0,length(sv))
  svoimsus = rep(0,length(sv))

  for (i in 1:length(sv)){

    d = sv[i]
    b_pv = rep(0,z)
    f_pv = rep(0,z)
    s_pv = rep(0,z)

    b = 0
    f = 0
    s = 0

    for (j in 1:z){
      x1 = rbinom(1,n,0.1)
      x2 = rbinom(1,n,d)

      b_pv[j] = barnard(x1,x2,n-
x1,n-x2,0.01)

      m = matrix(c(x1,n-x1,x2,n-
x2),nrow=2)

      f_pv[j] =
fisher.test(m,alternative="less")[1]

      s_pv[j] = g.test(m)
    }
  }
  for (k in 1:z){
    if (b_pv[k]<a){
      b = b + 1
    }
    if (f_pv[k]<a){
      f = f + 1
    }
    if (s_pv[k]<a){
      s = s + 1
    }
  }
  bvoimsus[i] = b/z
  fvoimsus[i] = f/z
  svoimsus[i] = s/z
}
bvoimsus
fvoimsus
```

```

        voimsused =
matrix(c(bvoimsus,fvoimsus,svoimsus,sv-0.1),ncol=4)
        return (voimsused)
}

v1 = voimsuseleidja1(100000,0.05,15)

plot(v1[,4],v1[,1],type="l",lwd=2,col="green",xlab=expression(rho),ylab="Võimus")
lines(v1[,4],v1[,2],col="blue",lwd = 2)
lines(v1[,4],v1[,3],lwd = 2)
abline(0.05,0,col="red")

voimsuseleidja2 = function(z,a,n){

    sv = seq(0.1,1,0.1)
    bvoimsus = rep(0,length(sv))
    fvoimsus = rep(0,length(sv))
    svoimsus = rep(0,length(sv))

    for (i in 1:length(sv)){

        d = sv[i]
        b_pv = rep(0,z)
        f_pv = rep(0,z)
        s_pv = rep(0,z)

        b = 0
        f = 0
        s = 0

        for (j in 1:z){
            n_b = rbinom(2,2*n,0.5)
            x1 = rbinom(1,n_b[1],0.1)
            x2 = rbinom(1,n_b[2],d)

            b_pv[j] =
barnard(x1,x2,n_b[1]-x1,n_b[2]-x2,0.01)

            m = matrix(c(x1,n_b[1]-
x1,x2,n_b[2]-x2),nrow=2)
            f_pv[j] =
fisher.test(m,alternative="less")[1]

            s_pv[j] = g.test(m)

        }
        for (k in 1:z){
            if (b_pv[k]<a){
                b = b + 1
            }
            if (f_pv[k]<a){
                f = f + 1
            }
            if (s_pv[k]<a){
                s = s + 1
            }
        }

        bvoimsus[i] = b/z
        fvoimsus[i] = f/z
        svoimsus[i] = s/z
    }
    voimsused =
matrix(c(bvoimsus,fvoimsus,svoimsus,sv-0.1),ncol=4)
    return (voimsused)
}

```

```

v2 = voimsuseleidja2(100000,0.05,15)

plot(v2[,4],v2[,1],type="l",xlab=expression(rho),ylab="Võimus",
col = "green", lwd = 2) #Barnardi joon
lines(v2[,4],v2[,2],col="blue", lwd = 2) #Fisheri
joon
lines(v2[,4],v2[,3], lwd = 2) #Toeparasuite joon
abline(0.05,0,col="red", type = "o")
voimsuseleidja3 = function(z,a,n){

    sv = seq(0,0.9,0.1)
    bvoimsus = rep(0,length(sv))
    fvoimsus = rep(0,length(sv))
    svoimsus = rep(0,length(sv))

    for (i in 1:length(sv)){

        d = sv[i]
        b_pv = rep(0,z)
        f_pv = rep(0,z)
        s_pv = rep(0,z)

        b = 0
        f = 0
        s = 0

        for (j in 1:z){
            x1 = rbinom(1,8,1-d)
            x2 = n-x1
            x3 = n-x1
            x4 = x1

            b_pv[j] =
barnard(x1,x2,x3,x4,0.01)

            m =
matrix(c(x1,x3,x2,x4),nrow=2)
            f_pv[j] =
fisher.test(m,alternative="less")[1]

            s_pv[j] = g.test(m)

        }
        for (k in 1:z){
            if (b_pv[k]<a){
                b = b + 1
            }
            if (f_pv[k]<a){
                f = f + 1
            }
            if (s_pv[k]<a){
                s = s + 1
            }
        }

        bvoimsus[i] = b/z
        fvoimsus[i] = f/z
        svoimsus[i] = s/z
    }
    voimsused =
matrix(c(bvoimsus,fvoimsus,svoimsus,sv-0.1),ncol=4)
    return (voimsused)
}

v3 = voimsuseleidja3(10000,0.05,15)

```

```

plot(v3[,4]+0.1,v3[,1],type="l",xlab=expression(rho),y=
b="Võimus",
col = "green", lwd = 2) #Barnardi joon
lines(v3[,4]+0.1,v3[,2],col="blue", lwd = 2) #Fisher
joon
lines(v3[,4]+0.1,v3[,3], lwd = 2) #Toeparasuhte
joon
abline(0.05,0,col="red")

```

---



---

```

# BARNARDI TESTSTATISTIKU JAOUTUS

```

```

b_ts_jaotus = function(n1,n2,pi){
  s = n1*n2
  waldid = rep(0,s)
  pr = rep(0,s)
  t = 0
  for (m in 0:(n1)){
    for (n in 0:(n2)){
      waldid[t] = Wald(m,n,n1,n2)
      t = t + 1
    }
  }

  k = length(waldid)
  y = rep(0,k)
  tt = 1
  for (i in waldid){
    m = 0
    for (j in 1:k){
      if (waldid[j]<i){
        m = m + 1
      }
    }
    y[tt] = m/k
    tt = tt + 1
  }
  plot(waldid,y,ylim=c(0.6,1),xlim=c(1.84,6), col
= "black",xlab="Waldi teststatistikud",ylab = "F(Waldi
testst.)")
  lines(c(1.84,1.84),c(0,1),col="red")
  #lines(c(-1.84,-1.84),c(0,1),col="red")
  return ("")
}

```

```

#FISHERI TESTSTATISTIKU JAOUTUS

```

```

f_ts_jaotus = function(n1,n2,k){
  waldid = rep(0,k)
  t = 0
  for (m in 0:(n1+1)){
    for (n in 0:(n2+1)){
      if (m + n == k){
        waldid[t] =
Wald(m,n,n1,n2)
        t = t + 1
      }
    }
  }
  print (waldid)
  k = length(waldid)
  y = rep(0,k)
  tt = 1
  for (i in waldid){
    m = 0
    for (j in 1:k){

```

```

      if (waldid[j]<=i){
        m = m + 1
      }
    }
    y[tt] = m/k
    tt = tt + 1
  }
  plot(waldid,y,ylim=c(0.6,1),xlim=c(1.8,6),col=
"blue",ylab="F(Waldi testst.)", xlab="Waldi
teststatistik")
  lines(c(1.84,1.84),c(0,1),col="red")
  return ("")
}

```

---



---

```

#Barnardi test 2 x 3 tabelile

```

```

binoom3 = function(xa,xb,xc,n1,n2,n3,pi){
  b = choose(n1,xa) * choose(n2,xb) *
choose(n3,xc) * pi^(xa+xb+xc) * (1-pi)^(n1 + n2 + n3 -
xa - xb - xc)
  return (b)
}

```

```

p3 = function(xa,xb,xc,xd,xe,xf,pi){
  n1 = xa + xd
  n2 = xb + xe
  n3 = xc + xf
  kriitP = binoom3(xa,xb,xc,n1,n2,n3,pi)
  summa = 0
  for (m in 0:(n1+1)){
    for (n in 0:(n2+1)){
      for (o in 0:(n3+1)){
        if (binoom3(m,n,o,n1,n2,n3,pi) <= kriitP){
          summa = summa +
binoom3(m,n,o,n1,n2,n3,pi)
        }
      }
    }
  }
  return (summa)
}

```

```

barnard3 = function(xa,xb,xc,xd,xe,xf,tapsus){
  x = seq(0,1,by=tapsus)
  #print("p-vaartus Barnardi yhepool testil")
  pi = max(p3(xa,xb,xc,xd,xe,xf,x))
  if (pi == "NaN"){
    pip = 1
  }
  else {
    pip = pi
    #b_voimsus(xa+xc,xb+xd,pip)
  }
  return(pip)
}

```

```

#2 x 3 tabelid ja testi kiiruse arvutused

```

```

multi3 = function(xa,xb,xc,arv1,arv2,arv3,n){
  NN = arv1 + arv2 + arv3
  return((choose(arv1,xa)*choose(arv2,xb)*choo
se(arv3,xc))/choose(NN,n))
}

```



```

# Fisher'i test 2 x 3 tabelile

fisher3 = function(xa,xb,xc,xd,xe,xf){
  n1 = xa + xd
  n2 = xb + xe
  n3 = xc + xf
  k = xa + xb + xc
  kriitP = multi3(xa,xb,xc,n1,n2,n3,k)
  Pvaartus = 0
  for (m in 0:n1){
    for (n in 0:n2){
      for (o in 0:n3){
        if (multi3(m,n,o,n1,n2,n3,k) <= kriitP &&
m+n+o == k){
          Pvaartus = Pvaartus +
multi3(m,n,o,n1,n2,n3,k)
        }
      }
    }
  }
  #print("Fisher'i tapne (2 x 3 tabel) uhepoolne p-
vaartus:")
}

```

```

return (Pvaartus)
}

#TESTI KIIRUSTE ARVUTUSED

b1 = seq(1,5,1)
f1 = seq(1,5,1)
b31 = seq(1,5,1)
f31 = seq(1,5,1)

for (i in seq(1,5,1)){
  b1[i] = system.time(barnard(10*i-
4,i+4,10*i+4,10*i-4,0.001))[3]
  f1[i] = system.time(fisher(10*i-
4,i+4,10*i+4,10*i-4))[3]
  b31[i] = system.time(barnard3(10*i-4,i+4,10*i-
7,10*i+4,10*i-4,10*i+7,0.001))[3]
  f31[i] = system.time(fisher3(10*i-4,i+4,10*i-
7,10*i+4,10*i-4,10*i+7))[3]
}

```

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

Mina Edwart Ždanovitš

(sünnikuupäev: 11.02.1991 )

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Barnardi täpse testi võrdlus Fisheri täpse testiga“, mille juhendaja on Märt Möls.
  - 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2.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, **06.05.2013**