Peatükk 3

Wilcoxoni astaksummatest (*Wilcoxon Rank-Sum Test*)

3.1 Teststatistiku konstrueerimine

Wilcoxoni astaksummatest on mitteparameetriline test kahe (sõltumatu) populatsiooni võrdlemiseks. Testprotseduuri toome sisse praktilist näidet kasutades.

Peame ravima viit gripihaiget. Soovime teada saada, kas erilise hiina tee abil saaksid haiged kiiremini terveks. Jagame viis haiget juhuslikult kahte rühma — ravi saavaks ja kontrollrühmaks (nemad saavad tavalist teed). Ravi saavasse rühma valime juhuslikult 3 patsienti, kontrollrühma jääb 2 patsienti. Jälgime patsiente ja märgime üles, kes kuna terveks saab. Esimesena terveks saanud patsiendi tähistame numbriga 1, teisena terveks saava haige tähistame numbriga 2 jne. Haigetele omistatud tervenemisaja astakud võivad jaguneda kahte gruppi 10-l erineval viisil ($C_3^5 = C_2^5 = 10$):

Ravi	1, 2, 3	1, 2, 4	1, 2, 5	1, 3, 4	1, 3, 5
Kontroll	4, 5	3, 5	3, 4	2, 5	2, 4

Ravi	1, 4, 5	2, 3, 4	2, 3, 5	2, 4, 5	3, 4, 5
Kontroll	2, 3	1, 5	1, 4	1, 3	1, 2

Juhul, kui ravil puuduks igasugune mõju oleks kõigi võimalike järjestuste esinemistõenäosus sama — kõik vaadeldud astakute jagunemised gruppidesse võiksid aset leida tõenäosusega 1/10. Kui aga ravi aitaks patsiente, siis kipuksid ravigrupis astakud olema väiksemad kui kontrollgrupis (ravigrupis saavad inimesed varem terveks). Siit tuleneb ka idee teststatistiku valikuks: liidame kõigi ravigruppi sattunud inimeste tervenemisaegade astakud. Kui saadav summa tuleb väga väike, siis vihjab see võimalusele, et ravist on kasu.

Formaalselt kirja pandult: olgu ravigrupis uuritava tunnuse väärtuste astakud $s_1, s_2, \ldots, s_n$ (ravi saanute rühmas oli kokku n inimest). Siis teststatistikuna kasutame suurust

$$W_s = s_1 + s_2 + ... + s_n.$$

Milline on statistiku W_s jaotus meie näite korral nullhüpoteesi kehtides? Leiame kõigi 10 variandi jaoks statistiku W_s väärtused ja saamegi kirja panna tema jaotuse nullhüpoteesi kehtides (võttes arvesse, et iga võimaliku kombinatsiooni esinemistõenäosus on $1/10$):

w	6	7	8	9	10	11	12
$P(W_s = w)$	0.1	0.1	0.2	0.2	0.2	0.1	0.1

Ennem uuringuandmete kogumist oleks tarvis määrata ka kasutatav olulisuse nivoo. Kuna tüüpilise olulisuse nivoo $0,05$ korral me nullhüpoteesi kummutada ei saakski (nii väikse valimi korral), siis valigem olulisuse nivoo $\alpha = 0,1$ (ja kasutame patuga pooleks ühepoolset hüpoteesi). Seega kui saaksime teststatistiku väärtuseks $W_s = 6$, siis me kummutaks nullhüpoteesi ja ütleks, et ravi aitab. Muul juhul jääksime nullhüpoteesi juurde.

Nüüd oleme teinud kõik otsused, mida tarvis teha enne katset ja võime asuda oma eksperimendi juurde. Oletame, et esimese, teise ja neljandana terveks saanud inimesed kuulusid ravitavate gruppi. Sellisel juhul $W_s = 1 + 2 + 4 = 7$, olulisustõenäosus $P(W_s \leq 7|H_0) = 0.1 + 0.1 = 0.2$ ja seega peame jääma nullhüpoteesi juurde.

Antud näites vaatlesime ravi saanud inimesterühma astakute summat. Sama hästi oleksime muidugi võinud vaadelda kontrollrühma kuuluvate patsientide tervenemisaegade astakute summat W_r. Paneme tähele, et statistikute W_s ja W_r väärtused on seotud lihtsa valemiga. Olgu kontrollrühmas tehtud m vaatlust ja ravirühmas n vaatlust, kokku mõlemas rühmas tehtud N vaatlust. Siis

$$W_s = 0.5N(N + 1) - W_r,$$

sest kõigi vaatluste astakute summa on $1 + 2 + ... + N = 0.5N(N + 1)$. Tänu sellisele üks-ühesele vastavusele on ükskõik, kumba rühma me loeme kontrollgrupiks. Tasub siiski täheldada, et statistikute W_s ja W_r jaotused nullhüpoteesi kehtides võivad olla erinevad. Statistiku W_s minimaalne võimalik väärtus on $0.5n(n+1)$, seevastu statistiku W_r minimaalne võimalik väärtus on $0.5m(m+1)$. See tekitab raskuseid tabelite koostamisel, ja sestap on kasutusele võetud ka transformeeritud statistik W_{xy}, mille puhul mainitud erinevus kaob. Statistiku W_{xy} leidmiseks tuleb Wilcoxoni teststatistikust maha lahutada tema minimaalne võimalik väärtus:

$$W_{xy} := W_s - 0.5n(n+1).$$

Seega ülaltoodud näite korral, kus $W_s = 7$, oleks $W_{xy} = 7 - 0.5 * 3 * (3 + 1) = 1$.

Teststatistikut W_{xy} tuntakse ka Mann-Whitney teststatistiku nime all ja temal baseeruvat testi kutsutase Mann-Whitney testiks. Pole raske näha, et Mann-Whitney testi otsused langevad kokku Wilcoxoni (astaksumma) testi omadega. Mann-Whitney teststatistikul on ka oma interpretatsioon. Nimelt võime vaadelda kõiki võimalikke paare (X, Y), kus X on pärit ühest populatsioonist võetud valimist (ravigrupist) ja Y on pärit teisest populatsioonist võetud valimist (kontrollgrupist). Mann-Whitney teststatistik näitab siis, kui paljudes vaadeldud paarides oli ravigruppi kuulunud patsient tulemus suurem (tervenemiseks läks kauem aega) kui kontrollgruppi kuulunud patsiendil,

$$W_{xy} = \#_{i,j}\{X_i > Y_j\}.$$

Kasutatavas näites oleksid kõikvõimalikud paarid järgmised:

Ravigrupi Astakud (X)	kontrollgrupi astakud (Y)	paarid	ravigrupi liige tervenes aeglasemalt
1	3	1 3	-
2	5	1 5	-
4		2 3	-
		2 5	-
		4 3	+
		4 5	-

Paaride arv mille puhul ravigruppi kuulunud patsiendil läks tervenemiseks kauem aega oli $1 = W_{xy}$.

Teoreem 3.1 $W_{xy} = \#_{i,j}\{X_i > Y_j\}$.

Tõestus.
Vaatleme ravirühma astakuid: ravirühma väikseima vaatluse astak $S_{(1)}$ näitab, mitu kontrollrühma patsienti on temast väiksemad + 1, st kui ravirühma

väikseim astak oleks 3 oleks kontrollrühmas 2 vaatlust, mis on temast väiksemad, seega antud vaatluse poolt moodustatud paaridest oleksid 2 sellised, kus kontrollrühma tulemus oleks olnud väiksem. Üldjuhul on siis pisemast ravirühma vaatlusest väiksemaid kontrollrühma vaatluseid $S_{(1)} - 1$ tükki.

Suuruselt järgmise ravirühma vaatluse astak on $S_{(2)}$, sellest vaatlusest väiksemaid kontrollrühma vaatluseid on aga $S_{(2)} - 2$ tükki (ses ravirühma üks vaatlustest oli ka sellest vaatlusest väiksem) jne. Seega:

$$\begin{aligned}
(S_{(1)} - 1) + (S_{(2)} - 2) + \ldots + (S_{(n)} - n) &= \#_{i,j}\{X_i > Y_j\} \\
W_s - 0.5n(n+1) &= \#_{i,j}\{X_i > Y_j\} \\
W_{xy} &= \#_{i,j}\{X_i > Y_j\}.
\end{aligned}$$

q.e.d.

Märkus: Lisaks W_{xy}-le võib kasutada ka statistikut W_{yx}-i:

$$W_{yx} := W_r - 0.5m(m+1) = \#_{i,j}\{Y_i > X_j\}.$$

3.2 Keskväärtus ja dispersioon H_0 kehtides

Alustuseks vaatame, milline on ühe vaatluse astakute jaotus ja kahe vaatluse astakute ühisjaotus H_0 kehtides.

Lemma 3.1 *Kui H_0 kehtib — populatsioonid on eristamatud — siis*

1. *i. vaatluse astaku S_i jaotus on*

$$P(S_i = k) = 1/N, k = 1, 2, ..., N.$$

2. *i. ja j. $(i, j = 1, 2, ..., N)$ vaatluse ühisjaotuseks on*

$$P(S_i = k, S_j = l) = \begin{cases} 1/[N(N-1)] & i \neq j \\ 0 & i = j \end{cases}.$$

Tõestus.

H_0 kehtides on kõik $N!$ vaatluste järjestust võrdvõimalikud. Neist $(N-1)!$ on sellised, kus $S_i = k$. Seega $P(S_i = k) = (N-1)!/N! = 1/N$. Selliseid järjestusi, kus $S_i = k, S_j = l$ on aga $(N-2)!$ tükki, seega

$$\begin{aligned}
P(S_i = k, S_j = l) &= (N-2)!/N! \\
&= 1/[N(N-1)].
\end{aligned}$$

Q.e.d.

Lemma 3.2 *Kui H_0 kehtib, siis:*

1. $ES_i = (N+1)/2$
2. $DS_i = (N^2-1)/12$
3. $Cov(S_i, S_j) = -(N+1)/12, i \neq j$

Tõestus.

$$\begin{aligned}
\mathrm{E}S_i &= 1/N(1+2+...+N) \\
&= 1/N * N(N+1)/2 \\
&= (N+1)/2
\end{aligned}$$

$$\begin{aligned}
\mathrm{D}Si &= 1/N(1^2+2^2+...+N^2) - (\mathrm{E}S_i)^2 \\
&= 1/N * N(N+1)(2N+1)/6 - (N+1)^2/4 \\
&= (4N^2+6N+2)/12 - (3N^2+6N+3)/12 \\
&= (N^2-1)/12.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(S_i, S_j) &= \mathrm{E}S_iS_j - \mathrm{E}S_i\mathrm{E}S_j \\
&= \frac{1}{N(N-1)}(1\cdot 1 + 1\cdot 2 + .. + 1\cdot N + 2\cdot 1 + \ldots + 2\cdot N + \ldots + N\cdot N - \\
&\quad -(1^2+2^2+\ldots+N^2)) - (N+1)^2/4 \\
&= \frac{1}{N(N-1)}\left(\frac{N(N+1)}{2}(1+2+...+N) - N(N+1)(2N+1)/6\right) - (N+1)^2/4 \\
&= \frac{1}{N(N-1)}(N(N+1)N(N+1)/4 - N(N+1)(2N+1)/6) - (N+1)^2/4 \\
&= \frac{1}{12N(N-1)}N(N+1)[3N(N+1) - 2(2N+1)] - (N+1)^2/4 \\
&= \frac{1}{12N(N-1)}N(N+1)(3N+2)(N-1) - (N+1)^2/4 \\
&= (N+1)(3N+2)/12 - (N+1)^2/4 \\
&= (N+1)(3N+2-3N-3)/12 \\
&= -(N+1)/12
\end{aligned}$$

Q.E.D.

Lemma 3.3 *Kui H_0 kehtib, siis*

1. $EW_s = n(N+1)/2$
2. $DW_s = nm(N+1)/12$

Tõestus

$$\begin{aligned} \mathrm{E}W_s &= \mathrm{E}(S_1 + S_2 + ... + S_n) \\ &= n(N+1)/2 \end{aligned}$$

$$\begin{aligned} \mathrm{D}W_s &= \mathrm{D}(S_1 + S_2 + \ldots + S_n) \\ &= n\mathrm{D}(S_i) + 2n(n-1)/2 \cdot \mathrm{Cov}(S_i, S_j) \\ &= n(N^2-1)/12 - n(n-1)(N+1)/12 \\ &= n(N+1)[N-1-(n-1)]/12 \\ &= nm(N+1)/12. \end{aligned}$$

Q.E.D.

Märkus. Mann-Whitney teststatistiku keskväärtus ja dispersioon on H_0 kehtides järgmised:

$$\begin{aligned} (\mathrm{E}W_{yx} =)\ \mathrm{E}W_{xy} &= mn/2 \\ (\mathrm{D}W_{yx} =)\ \mathrm{D}W_{xy} &= nm(N+1)/12. \end{aligned}$$

3.3 Asümptootiline jaotus

Suure katse puhul, kus n ja m on mõlemad suured, võib tõenäosuse $P(W_s > w)$ arvutamine kõigi võimaluste kokkulugemise teel osutuda väga töömahukaks. Sestap oleks hea, kui teaksime, millisele piirjaotusele läheneb meid huvitava statistiku W_s jaotus (nullhüpoteesi kehtides) suure valimi korral. Selgub, et selleks jaotuseks on normaaljaotus. Antud väite tõestus pole päris triviaalne, sest astakud pole sõltumatud ja seega on W_s sõltuvate juhuslike suuruste summa (tõestus pole ka lootusetult keeruline aga jääb siiski praegu välja). Teisisõnu öeldes, asümptootiliselt:

$$\frac{W_s - \mathrm{E}W_s}{\mathrm{D}W_s} \xrightarrow{\mathcal{L},H_0} \mathcal{N}(0,1).$$

Ehk, asendades keskväärtuse ja dispersiooni ülaltoodud valemisse, saame:

$$\frac{W_s - n(N+1)/2}{\sqrt{mn(N+1)/12}} \xrightarrow{\mathcal{L},H_0} \mathcal{N}(0,1).$$

Kontrollides ühepoolset hüpoteesi (teame eelnevalt, et töötlus ei mõju või kui mõjub, siis ainult uuritava tunnuse väärtuseid kahandavalt), saame Wilcoxoni testi olulisustõenäosuse leida suure valimi korral järgmise valemi abil:

$$\begin{aligned} P(W_s \le c|H_0) &= \Phi\left(\frac{c + \text{pidevuse parandus} - \mathrm{E}W_s}{\mathrm{D}W_s}\right) \\ &= \Phi\left(\frac{c + 0.5 - 0.5n(N+1)}{\sqrt{mn(N+1)/12}}\right) \end{aligned}$$

kus $\Phi(x)$ on standardse normaaljaotuse jaotusfunktsioon kohal x.

Kahepoolne hüpotees.

Oletame esmalt, et vaadeldavas valimis on astaksummatesti teststatistiku väärtus c väiksem kui nullhüpoteesi puhul oodatav testtatistiku väärtus, $c < \mathrm{E}W_s$ ehk $c < 0.5n(N+1)$.

Tõenäosus näha sedavõrd väikest või veel väiksemat teststatistiku väärtust on

$$\Phi\left(\frac{c + 0.5 - 0.5n(N+1)}{\sqrt{mn(N+1)/12}}\right).$$

Kahepoolse hüpoteesi puhul pakuvad aga huvi ka sedavõrd oodatavast suuremad või veel suuremad statistiku väärtused. Kui praegu on valimis nähtud teststatistiku väärtus $\mathrm{E}W_s - c = 0.5n(N+1) - c$ ühikut väiksem H_0 puhul oodatavast, siis kui tõenäoline on näha samavõrra H_0 puhul oodatavast väärtusest suuremat (või veel suuremat) teststatistiku väärtust? Vastav tõenäosus on leitav valemiga

$$
\begin{aligned}
P(W_s \geq \mathrm{E}W_s + (\mathrm{E}W_s - c)) = \\
&= P\left(W_s \geq n(N+1) - c\right) \\
&= 1 - P\left(W_s < n(N+1) - c\right) \\
&= 1 - \Phi\left(\frac{n(N+1) - c + \text{pidevuse korrektsioon} - \mathrm{E}W_s}{\sqrt{\mathrm{D}W_s}}\right) \\
&= 1 - \Phi\left(\frac{0.5n(N+1) - c - 0.5}{\sqrt{\mathrm{D}W_s}}\right),
\end{aligned}
$$

kust saame (kuna $1 - \Phi(x) = \Phi(-x)$):

$$
P\left(W_s \geq \mathrm{E}W_s + (\mathrm{E}W_s - c)\right) = \Phi\left(\frac{c + 0.5 - 0.5n(N+1)}{\sqrt{mn(N+1)/12}}\right).
$$

Liites mõlemad tõenäosused (sedavõrd väikese või veel väiksema teststatistiku saamise tõenäosuse ja samavõrra oodatavast suurema või veel suurema teststatistiku saamise tõenäosuse) saame olulisustõenäosuseks

$$
\text{p-väärtus} = 2 \cdot \Phi\left(\frac{c + 0.5 - 0.5n(N+1)}{\sqrt{mn(N+1)/12}}\right). \tag{3.1}
$$

Kui aga nähtud teststatistik peaks olema suurem kui nullhüpoteesi puhul oodatav värtus $\mathrm{E}W_s$, siis jõuame sarnase arutelu tulemusel järgmise p-väärtuseni:

$$
\text{p-väärtus} = 2 \cdot \Phi\left(-\frac{c + 0.5 - 0.5n(N+1)}{\sqrt{mn(N+1)/12}}\right). \tag{3.2}
$$

Valemite 3.1 ja 3.2 pealt võime kokku kirjutada igas situatsioonis kasutamiskõlbliku valemi kahepoolse testi p-väärtuse leidmiseks:

$$
\text{p-väärtus} = 2 \cdot \Phi\left(-\frac{|c + 0.5 - 0.5n(N+1)|}{\sqrt{mn(N+1)/12}}\right).
$$

3.4 Populatsiooni mudelist

Wilcoxoni astaksumma testi rakendamiseks pole meil tarvis endale ette kujutada kahte populatsiooni vms. Piisab sellest, et me jagame uuritavad juhuslikult kahte gruppi. Wilcoxoni testi puhul saame pärida, kas antud uuritavate indiviidide puhul üks töötlus on parem kui teine töötlus.

Enamasti aga soovime me teha järeldusi mitte meie valimisse kuuluvate indiviidide kohta, vaid mingi laiema populatsiooni kohta. Näiteks võime me võtta valimi patsientidest, ravida osasid ühel ja teisi teisel viisil ning lõpuks soovime öelda, milline ravi on parem antud haiguse patsientidele (kõigile patsientidele, mitte ainult meie valimisse sattunud inimeste jaoks). Või võtame kaks valimit kahest populatsioonist, $X \sim F(x)$ ja $Y \sim G(y)$ ning soovime teada, kas mõlemas populatsioonis on uuritava tunnuse jaotus samasugune, kas $F = G$?

Võtame valimi suurusega $N = n + m$. Kui nullhüpotees kehtib, $F = G$, siis on vaatluste $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ mistahes järjestus võrdvõimalik. Seega on tõenäosus näha esimeses valimis (X-id) astakuid $s_1, \ldots, s_n$ on võrdne võimaluste arvuga valida N vaatluse seast välja n vaatlust,

$$P(s_1, \ldots, s_n) = 1/\binom{N}{n}.$$

Seega jääb nullhüpoteesi kehtides Wilcoxoni astaksummatesti statistiku jaotus samasuguseks, ükskõik kas meil on tegemist juhuslike valimitega populatsioonidest või me kõigest jagame juhuslikult gruppidesse kogu meid huvitavat populatsiooni.

Mõnevõrra segadust võib tekkida siis, kui uuritava tunnuse jaotus pole (päriselt) pidev. Sellisel juhul on igas valimis erinev arv võrdseid vaatluseid ja Wilcoxoni astaksummatesti teststatistiku jaotus jääb sõltuma uuritavast jaotusest (sellest, millise tõenäosusega võime näha kokkulangevaid vaatluseid). Kuna aga uuritava tunnuse tegelik jaotus pole ju teada (milleks muidu me kasutame mitteparameetrilist testi), siis pole esmapilgul võimalik p-väärtust arvutada ja reaalselt statistilist testi teha. Lahenduseks on nn tingliku testi kasutamine — eeldame, et võrdsete astakute konfiguratsioon on antud, ja muretseme siis, et sellise konkreetse võrdsete astakute konfiguratsiooni puhul I-liiki viga ei tehtaks lubatust suurema tõenäosusega. Olulisustõenäosuse arvutamine toimub sellisel juhul täpselt samal viisil, kui tegime seda randomiseeritud katse korral. Kuna mistahes võrdsete astakute konfiguratsiooni puhul me esimest liiki viga me ei tee lubatust sagedamini, siis võime olla kindlad, et selline testprotseduur garanteerib, et I-liiki vea tegemise tõenäosus poleks suurem kui uurija poolt valitud olulisuse nivoo α.

3.5 Testi võimsus

Kas me suudame oma kasinate võimete ja võimaluste juures tõestada alternatiivse hüpoteesi (kui alternatiivne hüpotees ikka on õige)? Kui suurt valimit läheb tarvis, et suudaksime näidata erinevust populatsioonide vahel?

Üldjuhul käsitleb alternatiivne hüpotees ($F \neq G$) sedavõrd paljusid erinevaid situatsioone, et midagi üldist kõigi nende olukordade kohta öelda on raske. Kui uurijal on silma ees mingi konkreetne ettekujutus sellest, kuidas kahe populatsiooni jaotused võiksid teineteisest erineda, siis on sageli kõige lihtsam tee arvutisimulatsiooni abil hinnata testi võimsust (genereerida k sellist valimit, kus n vaatlust on võetud jaotusest F ja m vaatlust jaotusest G ja vaadata siis, kui paljudes genereeritud valimites suutis astaksummatest vastu võtta alternatiivse hüpoteesi).

Mida suuremaks muutuvad valimi suurused, seda raskemaks läheb ka nende genereerimine. Sestap vaatame alljärgnevalt mõnda võimalust leida astaksummatesti asümptootilist võimsust. Järgnevates arvutustes on veidi mugavam kasutada teststatistiku W_s asemel nn Mann-Whitney teststatistikut W_{xy}.

Alustame sellest, et leiame teststatistiku keskväärtuse ja dispersiooni alternatiivse hüpoteesi kehtides.

3.5.1 Keskväärtus ja dispersioon H_1 kehtides

Olgu antud m sõltumatut sama jaotusega juhuslikku suurust jaotusest F — $X_1, X_2, \ldots, X_m \sim F$ ja n s.s.j.j.s. jaotusest G — $Y_1, Y_2, \ldots, Y_n \sim G$.

Defineerime uued juhuslikud suurused U_{ij}:

$$U_{ij} = \begin{cases} 1 & \text{, kui } X_i < Y_j \\ 0 & \text{muidu} \end{cases} .$$

Juhusliku suuruse U_{ij} keskväärtuseks on

$$\mathrm{E}U_{ij} = P(X_i < Y_j) =: p_1.$$

Mann-Whitney teststatistik aga on esitatav juhuslike suuruste U_{ij} kaudu:

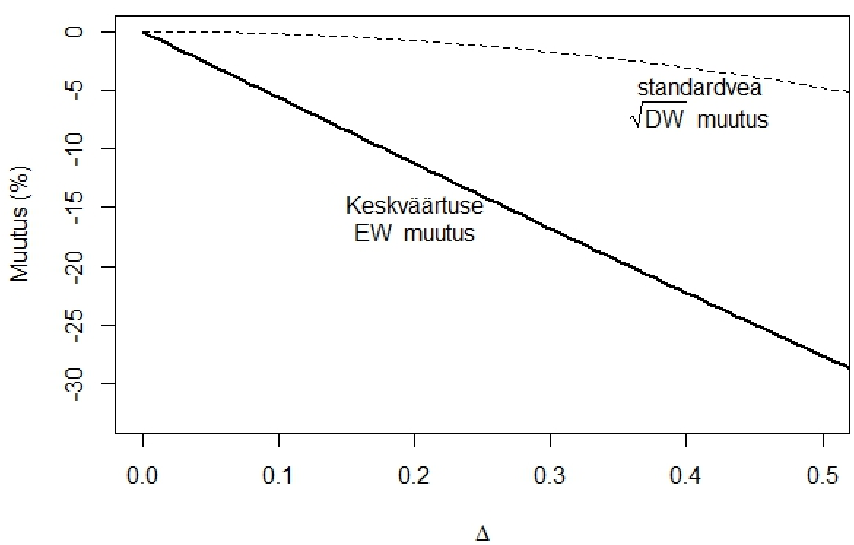
$$W_{xy} = \sum_i \sum_j U_{ij}.$$

Seega $\mathrm{E}W_{xy} = \sum_i \sum_j \mathrm{E}U_{ij} = mnp_1$.

Siit saame ühe täiendava võimaluse interpreteerida teststatistiku väärtust — nimelt on suurus $W_{xy}/(mn)$ nihketa hinnanguks tõenäosusele $p_1 = P(X < Y)$.

Mis juhtub teststatistiku dispersiooniga juhul, kui kehtib alternatiivne hüpotees? Kui erinevus jaotuste F ja G vahel on väike, siis muutub ka Mann-Whitney teststatistiku dispersioon (või standardviga) vaid veidi. Enamasti märkimisväärselt vähem kui teststatistiku keskväärtus. Kui näiteks võrreldavad populatsioonid oleksid normaaljaotusega, keskväärtuste erinevusega Δ, siis muutust Mann-Whitney teststatistiku keskväärtuses ja dispersioonis kirjedab joonis 3.1. Näeme, et väikeste Δ väärtuste puhul on muutus teststatistiku dispersioonis väga väike, ignoreeritav. Paljud autorid ütlevadki, et kui F ja G on lähedased, siis $\mathrm{D}W_{xy} \approx \mathrm{D}_{H0}W_{xy}$. Ka antud konspektis on hiljem kasutatud seda ligikaudset võrdust. Siinkohal paneme aga kirja ka valemi teststatistiku dispersiooni täpseks arvutamiseks.

Joonis 3.1: Teststatistiku W_{xy} keskväärtuse ja standardvea protsentuaalne muutus, kui keskväärtuste erinevus on Δ ja uuritava tunnuse jaotuseks on normaaljaotus.



$$\mathrm{D}W_{xy} = \sum_i \sum_j \mathrm{D}U_{ij} + \sum_i \sum_j \sum_k \sum_l \mathrm{Cov}(U_{ij}, U_{kl}),$$

kus kovariatsioonide summeerimine toimub üle nende i, j, k ja l väärtuste, mille puhul kas $i \neq k$ ja/või $j \neq l$ (variandi $i = k$ ja $j = l$ puhul on tegemist

U_{ij} dispersiooniga).

Antud summas on kolme erinevat tüüpi kovariatsioone. Kui $i \neq k$ ja $j \neq l$, siis leitakse U_{ij} ja U_{kl} täiesti erinevaid X ja Y tunnuse väärtuseid kasutades ja on sõltumatud (kuna on sõltumatute juhuslike suuruste funktsioonid). Seega on vastavad kovariatsioonid kõik nullid.

Juhtumeid, kus $i = k$ aga $j \neq l$ on $mn(n-1)$ tükki (X-e on m tükki, j valikuks n võimalust Y-te seast ja k valikuks jääb järgi $n-1$ võimalust). Liidetavaid, kus $i \neq k$ aga $j = l$ on aga $nm(m-1)$ tükki. Seega

$$\mathrm{D}W_{xy} = mn\mathrm{D}U_{ij} + mn(n-1)\mathrm{Cov}(U_{ij}, U_{il}) + nm(m-1)\mathrm{Cov}(U_{ij}, U_{kj}). \quad (3.3)$$

Leiame valemis kirjas olevad dispersioonid ja kovariatsioonid. Kuna U_{ij} on binaarne, siis

$$\mathrm{D}U_{ij} = p_1(1-p_1).$$

Kuna $\mathrm{E}U_{ij}U_{il} = P(U_{ij} = U_{il} = 1) = P(X_i < Y_j, X_i < Y_l) =: p_2$ siis

$$\begin{aligned} \mathrm{Cov}(U_{ij}, U_{il}) &= \mathrm{E}U_{ij}U_{il} - \mathrm{E}U_{ij}\mathrm{E}U_{il} \\ &= p_2 - p_1^2. \end{aligned}$$

Tähistades $p_3 := P(X_i < Y_j, X_k < Y_j)$ saame analoogselt, et $\mathrm{Cov}(U_{ij}, U_{kj}) = p_3 - p_1^2$.

Asendades saadud tulemused valemisse 3.3 saame

$$\mathrm{D}W_{xy} = mnp_1(1-p_1) + mn(n-1)(p_2 - p_1^2) + nm(m-1)(p_3 - p_1^2).$$

Juhul kui H_0 kehtib, siis $p_1 := P(X < Y) = 1/2$ ja $p_2 = p_3 = 1/3$ (juhuslikult valitud vaatluse tõenäosus olla väikseim kolmest sama jaotusega vaatlusest). Sellisel juhul aga $\mathrm{E}W_{xy} = mnp_1 = mn/2$ ja

$$\begin{aligned} \mathrm{D}W_{xy} &= mn/4 + mn(n-1)(1/3 - 1/4) + nm(m-1)(1/3 - 1/4) \\ &= mn/12(3 + n - 1 + m - 1) \\ &= mn/12(n + m + 1) \\ &= mn(N+1)/12. \end{aligned}$$

Saadud tulemused langevad kokku varemnähtud tulemustega.

3.5.2 Asümptootiline võimsus

Kontrollime nullhüpoteesi $H_0 : G = F$ kehtivust. Oletame, et nullhüpotees lükatakse ümber, kui teststatistik osutub suuremaks kriitilisest väärtusest

c_{krit}, $W_{xy} \geq c_{krit}$ (vaatame ühepoolset hüpoteesi). Siis alternatiivse hüpoteesi ($F \neq G$) kehtides on testi võimsus $\Pi(F,G)$ leitav valemist:

$$\begin{aligned}
\Pi(F,G) &= P(W_{xy} \geq c_{krit}) \\
&= P\left(\frac{W_{xy} - mnp_1}{\sqrt{\mathrm{D}W_{xy}}} \geq \frac{c_{krit} - mnp_1}{\sqrt{\mathrm{D}W_{xy}}}\right) \\
&\approx 1 - \Phi\left(\frac{c_{krit} - mnp_1}{\sqrt{\mathrm{D}W_{xy}}}\right), \qquad (3.4)
\end{aligned}$$

sest

$$\frac{W_{xy} - mnp_1}{\sqrt{\mathrm{D}W_{xy}}} \to N(0,1).$$

Teades p_1, p_2 ja p_3 väärtuseid, saame leida testi võimsuse. Nimetatud tõenäosuste leidmine on vahel võimalik, aga sageli on just p_2 ja p_3 jaoks millegi mõistliku väljapakkumine praktikas keerukas. Sestap proovime leida lisaeeldusi ja täiendavaid lihtsustusi tehes praktikas lihtsamini kasutatava valemi.

Esmalt leiame, milline on kriitiline väärtus c_{krit}:

$$\begin{aligned}
\alpha &= P(W_{xy} \geq c_{krit} | H_0) \\
\alpha &= P\left(\frac{W_{xy} - nm/2}{\sqrt{\mathrm{D}_{H_0} W_{xy}}} \geq \frac{c_{krit} - nm/2}{\sqrt{\mathrm{D}_{H_0} W_{xy}}} | H_0\right) \\
\alpha &\approx 1 - \Phi\left(\frac{c_{krit} - nm/2}{\sqrt{\mathrm{D}_{H_0} W_{xy}}}\right) \\
z_{1-\alpha} &\approx \frac{c_{krit} - nm/2}{\sqrt{\mathrm{D}_{H_0} W_{xy}}}
\end{aligned}$$

Kus $z_{1-\alpha}$ on standardse normaaljaotuse $1-\alpha$-kvantiil. Paari lihtsa teisendusega (ja arvestades, et $\mathrm{D}_{H_0} W_{xy} = mn(N+1)/12$) jõuame soovitud tulemuseni:

$$c_{krit} = z_{1-\alpha}\sqrt{mn(N+1)/12} + nm/2.$$

Asendades saadud tulemuse võimsuse valemisse 3.4 saame:

$$\begin{aligned}
\Pi(F,G) &= 1 - \Phi\left(\frac{z_{1-\alpha}\sqrt{mn(N+1)/12} + nm/2 - mnp_1}{\sqrt{\mathrm{D}W_{xy}}}\right) \\
&= 1 - \Phi\left(\frac{nm(0.5 - p_1) - z_\alpha\sqrt{mn(N+1)/12}}{\sqrt{\mathrm{D}W_{xy}}}\right). \qquad (3.5)
\end{aligned}$$

Saadud valemit on võimalik mõnikord tõepoolest kasutada. Näiteks eeldades, et uuritava tunnuse jaotus mõlemas populatsioonis on normaaljaotus, sama hajuvusega (standardhälbega σ) ning eeldades, et keskväärtuste erinevus populatsioonide vahel on Δ, saame tõenäosused p_1, p_2, p_3 ja ühepoolse testi võimsuse leida näiteks järgmise R programmi abil:

```
# Valimite suurused
n=30; m=10

# Uuritava tunnuse hajuvus (mõlemas populatsioonis)
sigma=1

# Keskväärtuste erinevus
delta=0.5

# Kasutatav olulisuse nivoo
alpha=0.05

# Leiame tõenäosused p1, p2, p3 (juhul, kui on normaaljaotus)
library(mvtnorm)
p1=pnorm(0, -delta, sd=sqrt(2*sigma**2))
p2=pmvnorm( lower=c(-Inf,-Inf), upper=c(0,0),
    mean=c(-delta, -delta),
    sigma=cbind(c(2*sigma**2,sigma**2), c(sigma**2,2*sigma**2)))

# Teststatistiku dispersioon ja keskväärtus H1 kehtides:
DW=m*n*p1*(1-p1)+m*n*(n-1)*(p2-p1**2)+ n*m*(m-1)*(p3-p1**2)
EW=m*n*p1

# Leiame ühepoolse testi võimsuse
1-pnorm((n*m*(0.5-p1)-qnorm(alpha)*sqrt(m*n*(m+n+1)/12))/sqrt(DW))
```

Toodud valemit otse kasutada on enamasti siiski raske. Paremal juhul võib eeldada, et $\mathrm{D}W_{xy} \approx \mathrm{D}_{H0}W_{xy}$, millisel juhul läheb võimsusarvutuse jaoks tarvis vaid suurust $p_1 := P(X < Y)$. Ka selle ühe tõenäosuse leidmine ei pruugi praktikas olla kuigi lihtne ülesanne. Milline võiks näiteks olla tõenäosus, et hommikuti putru söönud laps on pikem putru mittesöönud lapsest? Palju lihtsam on aga saada vanaemalt kätte eksperthinnang: tänu pudrule kasvab laps 5cm pikemaks. Vaatamegi alljärgnevalt, kuidas kasutada valemit 3.5 juhul, kui teame testitava töötluse kohta, palju ta uuritava tunnuse väärtust muudab.

Erinegu kaks populatsiooni teineteisest vaid nihke poolest, st $Y = X + \Delta$ ehk $F_X(x) = G_Y(x + \Delta)$. Sellisel juhul $p_1 = P(X - Y < 0) = P(X - (Y - \Delta) < \Delta)$. Kuna nii X kui ka $Y - \Delta$ on sama jaotusega, siis $p_1 = F^*(\Delta)$, kus F^* on kahe F jaotusega juhusliku suuruse vahe jaotus.

Funktsiooni $f(x)$ Taylori rida kohal a on kujul

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

Tähistades F^* tihedusfunktsiooni f^*-ga saame kirja panna funktsiooni F^* Taylori rittaarenduse kohal 0:

$$\begin{aligned} F^*(\Delta) &= F^*(0) + f^*(0)\Delta + \frac{(f^*)'(0)}{2}\Delta^2 + \ldots \\ &\approx F^*(0) + f^*(0)\Delta \end{aligned}$$

Kuna $F^*(0) = 0.5$ — sest kahe sama jaotusega juhusliku suuruse vahe mediaan (ja keskväärtus) on 0 — siis saame, et

$$p_1 - 0.5 \approx \Delta f^*(0).$$

Asendame saadud tulemuse valemisse 3.5 ja saame:

$$\begin{aligned} \Pi(F, G) &= 1 - \Phi\left(\frac{-mn\Delta f^*(0) - z_\alpha\sqrt{\frac{1}{12}mn(N+1)}}{\sqrt{\mathrm{D}W_{xy}}}\right) \\ &= \Phi\left(\frac{mn\Delta f^*(0) + z_\alpha\sqrt{\frac{1}{12}mn(N+1)}}{\sqrt{\mathrm{D}W_{xy}}}\right), \end{aligned}$$

sest $1 - \Phi(x) = \Phi(-x)$. Kui Δ on väike, siis $\mathrm{D}W_{xy} \approx \frac{1}{12}mn(N+1)$ ja

$$\Pi(F, G) = \Phi\left(\sqrt{\frac{12mn}{N+1}}\Delta f^*(0) + z_\alpha\right).$$

Näide 3.1 *Küsitlusfirmas töötab 20 ankeedisisestajat kahes tööruumis. Meid huvitab, kas taustamuusika mängimine tööruumis võiks tõsta nende töö effektiivsust. Teame varasemast, et ühe töötaja poolt päeva jooksul sisestatud lehekülgede arvu dispersioon on 32. Kui suure tõenäosusega suudaksime tõestada olulisuse nivool 0,05 taustamuusikast tingitud muutuse (eeldame hetkel, et meid huvitab vaid ühepoolne hüpotees)? Oletame, et kui muutus eksisteerib,* *siis tõstab taustamuusika sisestatud lehekülgede arvu 5 võrra. Soovime kasutada Wilcoxoni testi (sest me ei tea kindlalt päeva jooksul sisestatud lehekülgede arvu jaotust), aga võimsusarvutust tehes eeldagem, et vaadeldud tunnus on normaaljaotusega.*

Lahendus.
Kui $X, X' \sim N(\mu, \sigma^2 = 32)$ *siis* $X - X' \sim N(0, 2\cdot\sigma^2)$ *ja seega* $F^* = N(0, 2\cdot\sigma^2)$ *ning*

$$\begin{aligned} f^*(x) &= \frac{1}{\sqrt{2\pi(2*\sigma^2)}} \exp\left(-\frac{(x-\mu)^2}{2\sqrt{(2*\sigma^2)}}\right) \\ f^*(0) &= \frac{1}{2\sigma\sqrt{\pi}}. \end{aligned}$$

Seega oleks Wilcoxoni astaksummatesti võimsus $\Delta = 5$ *korral*

$$\begin{aligned} \Pi(F,G) &= \Phi\left(\sqrt{\frac{12mn}{N+1}}\frac{\Delta}{2\sigma\sqrt{\pi}} + z_\alpha\right) \\ &= \Phi\left(\sqrt{\frac{12\cdot 10\cdot 10}{20+1}}\frac{5}{2\sqrt{32}\sqrt{\pi}} - 1,64\right) \\ &= \Phi\left(0,24\right) \\ &= 0,5948. \end{aligned}$$

Testi võimsus 0,59 on madalapoolne, aga riskialdima eksperimentaatori jaoks ehk siiski piisav katsega alustamiseks. Kui leiaksime asümptootilise võimsuse abil testi täpse võimsuse, siis see oleks 0,55 (valimimaht 10+10 pole veel täiesti piisav asümptootika probleemivabaks tööks, aga väga viltu nüüd ka ei läinud...).

Vahel soovitakse testi võimsuse asemel leida hoopis valimi suurust, mida oleks vaja teatud võimsuse saavutamiseks. Valimi mahu leidmiseks, mis tagaks testi võimsuse β, peaksime leidma m ja n väärtused järgmisest võrrandist:

$$\beta = \Phi\left(\sqrt{\frac{12mn}{N+1}}\Delta f^*(0) + z_\alpha\right).$$

Saamaks ühest lahendit tehakse mingi eeldus m ja n soovitava suhte kohta (võib fikseerida ka ühe neist ja küsida, kui suure valimi peaksime võtma teisest populatsioonist...). Näiteks võime otsustada, et $m = n$. Sellisel juhul (ja arvestades, et $1/(N+1) \approx 1/N$ suure valimi korral):

$$\begin{aligned}
\beta &= \Phi\left(\sqrt{\frac{12mn}{N+1}}\Delta f^*(0) + z_\alpha\right) \\
z_\beta &= \sqrt{\frac{12mn}{N+1}}\Delta f^*(0) + z_\alpha \\
z_\beta - z_\alpha &= \sqrt{\frac{12n^2}{2n}}\Delta f^*(0) \\
\frac{z_\beta - z_\alpha}{\Delta f^*(0)} &= \sqrt{6n} \\
\frac{(z_\beta - z_\alpha)^2}{\Delta^2 (f^*(0))^2} &= 6n \\
n &= \frac{(z_\beta - z_\alpha)^2}{6\Delta^2 (f^*(0))^2}
\end{aligned}$$

Näide 3.2 *Pöördume tagasi eelmise, ankeedisisestajate näite juurde. Vaatame, kui suurt valimit läheks meil vaja, et saavutada 90% võimsust:*

$$\begin{aligned}
n &= \frac{(z_\beta - z_\alpha)^2}{6\Delta^2 \frac{1}{4\sigma^2\pi}} \\
&= (z_\beta - z_\alpha)^2 \frac{\sigma^2}{\Delta^2} 2\pi/3 \\
&= (1,28\ldots - (-1,64\ldots))^2 \cdot \frac{32}{5^2} \cdot 2 \cdot 3,14\ldots/3 \\
&= 22,95.
\end{aligned}$$

Mõlemasse katserühma vajame seega 23 ankeedisisestajat.

3.5.3 Wilcoxoni ja t-testi võrdlus

Paljudel tekib kindlasti küsimus — kui palju me võidaksime testi võimsuses, kui kasutaksime Wilcoxoni testi asemel t-testi? Erinevate testide võrdlemisel kasutatakse sageli nn *Pitman efficiency*'t ehk Pitmani effektiivsust (vahel kutsutakse ka relatiivseks efektiivsuseks ehk *relative efficiency*. Nimetatud näitaja leidmiseks tuleb välja selgitada, kui suurt valimit läheb vaja ühe testi korral ja kui suurt valimit teise testi korral et saavutada etteantud võimsus. Kahe valimi suuruse suhe ongi Pitmani effektiivsus. Näiteks Wilcoxoni ja t-testi puhul oleks Pitmani effektiivsus $e_{t,W}$ leitav suhtena

$$e_{t,W} = \frac{n_t}{n_{Wilcoxon}},$$

kus n_t oleks valimi suurus, mida t-test vajab võimsuse Φ saavutamiseks ja $n_{Wilcoxon}$ on Wilcoxoni testi poolt vajatav valimi suurus sama testi võimsuse korral. Arvutame t-testi suhtelise efektiivsuse Wilcoxoni testi suhtes suurte valimite jaoks (kus võime kasutada teststatistikute asümptootilisi jaotusi). Asümptootilise Relatiivse Efektiivsuse tähistamiseks kasutatakse sageli lühendit ARE.

Sarnast tuletuskäiku kui Wilcoxoni testi puhul kasutasime võime jõuda t-testi puhul vajamineva valimi mahu suuruse arvutamisel tulemuseni

$$n_t = 2(z_\beta - z_\alpha)^2 \frac{\sigma^2}{\Delta^2}.$$

Juhul, kui vaatlused oleksid normaaljaotusega, siis oleks t-testi ARE Wilcoxoni astaksumma testi suhtes seega:

$$\begin{aligned} e_{t,W} &= \frac{2(z_\beta - z_\alpha)^2 \cdot \sigma^2/\Delta^2}{(z_\beta - z_\alpha)^2 \cdot \sigma^2/\Delta^2 \cdot \frac{2\pi}{3}} \\ &= \frac{3}{\pi} \\ &\approx 0,955 \end{aligned}$$

Seega justnagu "kaotaksime" Wilcoxoni testi kasutades umbes 5% oma valimi mahust. Samas võime teha sarnaseid arvutusi ka teiste jaotuste korral. Kui valim on suur, võib ju t-testi kasutada ka siis, kui vaatlused pole normaaljaotusega. Alljärgnevalt mõned ARE-numbrid erinevate jaotuste puhul, nihkemudel:

normaaljaotus	logistiline	eksponentjaotus	double exponent
0,955	1,097	3	1,5

Enamike mittenormaalsete jaotuste puhul on Wilcoxoni test tundlikum ja võib vahel vajada isegi 3-4 korda väiksemat valimit kui t-test. Eksisteerib ka alampiir — t-testi ja Wilcoxoni testi relatiivne efektiivsus ei saa (nihkemudeli korral suvalise pideva jaotuse F puhul) olla väiksem kui 0,864 ($e_{t,W} \geq 0,864$). Vaata ka joonist 3.2. Vastavat ülempiiri ei eksisteeri — Wilcoxoni test võib olla kuidas palju tundlikum t-testist. Seega võimalik võit Wilcoxoni astaksummatesti kasutades võib olla väga suur, aga kaotus (kui vaatlused näiteks tõepoolest on normaaljaotusega) üsnagi pisike.

Märkus: samasugune alampiir kehtib ka Wilcoxoni astakmärgitesti korral (võrreldes sõltuvate valimite jaoks mõeldud t-testiga).