Peatükk 2

Wilcoxoni astakmärgitest

2.1 Motivatsioon ja teststatistik

Wilcoxoni astakmärgitesti (*Wilcoxon Signed-Rank Test*) kasutatakse kahe sõltuva valimi võrdlemiseks. Oletame näiteks, et soovime võrrelda, millise väetise abil saame parema saagi. Iga katsepõld jagatakse kaheks pooleks, millest ühte poolt väetatakse ühe väetisega, teist poolt aga teise väetisega. Sügisel vaadatakse kuidas saagiga lood on. Näiteks võis vaatluse all olla kolm põldu ja olgu saadud saagid järgmised:

	I pool	II pool	erinevus
1. põld	76	78	2
2. põld	82	91	9
3. põld	80	86	6

Kahel põllul oli parem uue väetisega väetatud pool, ühel põllul andis parema tulemuse vana väetis. Kas see on ükskõik, millisel põllul oli vana väetis parem? Vaatame kahte hüpoteetilist olukorda. Ühel juhul on uus väetis paaril põllul mäekõrguselt parem vanast (6 ja 9 ühikut) ja ühel põllul oli vana väetisega väetatud pool napilt parem (2 ühikut). Teisel juhul annab uue väetisega väetatud pool kahel põllul napilt paremat saaki (2 ja 6 ühikut), üks vana väetisega väetatud põllupool on aga märkimisväärselt parem kui uue väetisega väetatud põllupool (9 ühikut). Kas mõlemal kirjeldatud juhul on andmed uue väetise kasulikkusest sama tugevad? Või arvame, et esimesena kirjeldatud juhul usume meelsamini uue väetise paremusse kui teisena kirjeldatud juhu puhul?

Loomulikult võime taoliste andmete puhul kasutada märgitesti. Leiame iga põllu jaoks saagikuste erinevuse (uue väetisega poole saak - vana väetisega poole saak), tähistame saadud saagikuste erinevuse i. põllul sümboliga X_i. Võime kontrollida märgitesti abil, kasutades teststatistikut $S(X,0) = \sum_{i=1}^{n} \mathrm{sgn}(X_i)$, kas erinevuste mediaan võiks olla null.

Aga võime ka jälgida oma sisetunnet ja otsida midagi märgitestist paremat. Sellist, mis arvestaks mingil moel ka saagikuste erinevuse suurust. Wilcoxoni astakmärgitesti statistik võtabki arvesse lisaks erinevuse märgile ka erinevuse (absoluutväärtuse) suurust (astakut):

$$W = \sum_{i=1}^{n} \mathrm{sgn}(X_i)\mathrm{rank}(|X_i|).$$

2.2 Statistiku W jaotus nullhüpoteesi kehtides

Kui uus ja vana väetis on võrdväärselt head, siis tulenevad erinevused saakides lihtsalt põllupoolte erinevusest. Kumba põllupoolt väetatakse uue väetisega, on heas katses valitud juhuslikult (otsustatakse mündiviske abil vms). Kui vaatame saagikuste erinevusi — saak uue väetiega miinus saak vana väetisega — siis nullhüpoteesi kehtides peaksid märgid erinevuste ees olema täiesti juhuslikud. Antud näites on vaadeldud kolme põldu, seega on võimalikke märgikombinatsioone erinevuste ees $2^3 = 8$, ning järgnevad katsetulemused (valimid) peaksid kõik esinema võrdse tõenäosusega:

erinevused (x_i)			$\mathrm{sgn}(x_i)\mathrm{rank}(\|x_i\|)$			$W(x_i)$	$P(W = W(x_i))$
-2	-6	-9	-1	-2	-3	$W = -6$	1/8
$+2$	-6	-9	$+1$	-2	-3	$W = -4$	1/8
-2	$+6$	-9	-1	$+2$	-3	$W = -2$	1/8
$+2$	$+6$	-9	$+1$	$+2$	-3	$W = 0$	1/8
-2	-6	$+9$	-1	-2	$+3$	$W = 0$	1/8
$+2$	-6	$+9$	$+1$	-2	$+3$	$W = +2$	1/8
-2	$+6$	$+9$	-1	$+2$	$+3$	$W = +4$	1/8
$+2$	$+6$	$+9$	$+1$	$+2$	$+3$	$W = +6$	1/8

Ülaltoodud tabelit kasutades võime välja kirjutada Wilcoxoni astakmärgitesti statistiku W jaotuse juhul, kui $n = 3$, vaata tabelit 2.1.

Oletame nüüd, et meie katses tulid erinevused 6 ja 9 uue väetise kasuks ning erinevus 2 vana väetise kasuks. Seega $W = -4$. Leiame olulisustõenäosuse (p-value) — milline on tõenäosus näha sedavõrd ekstremaalset või veel ekstremaalsemat statistiku väärtust nullhüpoteesi kehtides? Ühepoolse hüpoteesi korral, kui uus väetis saab olla vaid kas samahea või parem

Tabel 2.1: Statistiku W jaotus nullhüpoteesi kehtides, $n = 3$

w	-6	-4	-2	0	2	4	6
P(W=w)	1/8	1/8	1/8	2/8	1/8	1/8	1/8

kui vana väetis, oleksid sama ekstreemse ($W = -4$) või veel ekstreemsema ($W = -6$) statistiku väärtuse saamise tõenäosus nullhüpoteesi kehtides $1/8 + 1/8 = 0,25$, seega olulisustõenäosus on ühepoolse hüpoteesi korral 0,25. Kahepoolse hüpoteesi korral (kui uus väetis võib põhimõtteliselt olla ka kehvem kui vana) on sama ekstreemse või veel ekstreemsemate statistiku väärtuste hulk $\{-6, -4, 4, 6\}$ suurem ja testi olulisustõenäosuseks tuleb 0,25*2=0,5. Seega antud juhul ei suuda me alternatiivset hüpoteesi tõestatuks lugeda mistahes mõistlikku olulisuse nivood kasutades.

Võime ka tähele panna, et kolme vaatluse korral poleks ühegi katsetulemuse korral võimalik tõestada uue väetise paremust vanast olulisuse nivool 0,05!

Seega üldjuhul on võimalik Wilcoxoni astakmärgitesti olulisustõenäosust leida näiteks nii: vaatame läbi 2^n erinevat märgikombinatsiooni astakute ees, leiame iga kombinatsiooni jaoks statistiku W väärtuse ning kirjutame välja W jaotuse juhul, kui nullhüpotees kehtib. Saadud jaotust kasutades on juba lihtne leida testi olulisustõenäosust. Tasub meeles pidada, et iga valimi suuruse n korral tuleb statistiku W jaotus erinev!

Samas pole olulisustõenäosuse leidmiseks tarvis tervet jaotust välja kirjutada. Piisab, kui loeme kokku ainult nähtud W väärtuse ja temast veel ekstreemsemate väärtuste saamise tõenäosused nullhüpoteesi kehtides.

Vaatame järgmist näidet

Näide 2.1 *Ravimifirma soovis uurida, kas tema poolt välja töötatud spetsiaalset haavaplaastrit kasutades paranevad haavad kiiremini kui traditsioonilist ravimeetodit (haavade kokkuõmblemist) kasutades. Uuringusse värvati 10 katsealust (rotti) ja igale rotile tehti kaks samapikka sisselõiget. Üks haavadest õmmeldi kokku (kasutati traditsioonilist raviviisi) ja teine plaasterdati kinni. Lasti siis haavadel paar päeva paraneda. Haava paranemise headust mõõdeti paari päeva pärast järgmiselt: vaadati, kui tugevat jõudu läheb vaja haava uuesti lahti rebimiseks (seega suuremad numbrid näitavad paremini paranenud haavu). Katsetulemused on toodud tabelis 2.2.*

Leiame kogutud andmete põhjal Wilcoxoni astakmärgitesti statistiku vää-

Tabel 2.2: Kliiniline katse parima haavaravi leidmiseks

rott	plaaster	õmblus	plaaster-õmblus	$sgn(x_i)rank(\|x_i\|)$
1	659	452	207	6
2	984	587	397	9
3	397	460	-63	-1
4	574	787	-213	-7
5	447	351	96	4
6	479	277	202	5
7	676	234	442	10
8	761	516	245	8
9	647	577	70	3
10	577	513	64	2

tuse:

$$W = 6 + 9 - 1 - 7 + 4 + 5 + 10 + 8 + 3 + 2 (= 39).$$

Antud juhul on erinevaid võimalusi astakute ette märke panna $2^{10} = 1024$*, kõiki neid läbi vaadata on raske. Aga vaatame, mitu märgikombinatsiooni on sellised, mille puhul tuleks teststatistiku väärtus 39 või suurem. Astakud, mille ette miinusmärgi pannes saame sama ekstreemse või veel ekstreemsema statistiku väärtuse, on ära toodud alljärgnevas tabelis:*

miinusmärkide arv	*astakud*	*juhte kokku*
0	$\emptyset$	*1*
1	*1, 2, 3, 4, 5, 6, 7, 8*	*8*
2	*(1;2), (1;3),(1;4),(1;5),(1;6),(1;7), (2;3),(2;4),(2;5),(2;6),(3;4),(3;5)*	*12*
3	*(1;2;3), (1;2;4), (1;2;5), (1;3;4)*	*4*

Kokku on seega neist 1024-st juhust selliseid, mille puhul W *väärtus tuleks sama ekstreemne või veelgi ekstreemsem kui vaadeldud statistiku väärtus* $1 + 8 + 12 + 4 = 25$*. Kui me teaksime ette (juba enne andmete kogumist/nägemist), et plaaster saab õmblusest vaid parem olla (mitte halvem), võiksime kontrollida ühepoolset hüpoteesi. Sellisel juhul oleks sedavõrd ekstreemse või veel ekstreemsema statistiku väärtuse nägemise tõenäosus nullhüpoteesi kehtides (p-value)* $25/1024 = 0.0244\ldots$*. Kui me ei saa kindlalt öelda, et "plaastriga haavad ei saa paraneda kehvemini kui õmblust kasutades", peaksime kasutama kahepoolset hüpoteesi. Sel juhul saaksime olulisustõenäosuseks* $0.0244 * 2 = 0,0488$*. Antud näite korral saame seega tõestatuks lugeda (olulisuse nivool 0,05), et plaaster parandab haavu paremini.*

Juhul kui vaatluste arv läheb suureks, võib täpse olulisustõenäosuse leidmine isegi arvuti abil muutuda keeruliseks (näiteks täiesti mõistliku valimi suuruse $n = 100$ korral peaksime läbi vaatama $\approx 10^{30}$ erinevat märgikombinatsiooni...). Isegi siis, kui peaksime olulisustõenäosuse leidmiseks läbi vaatama vaid umbes 5% neist, oleks seda tänapäeva arvutite jaoks liiga palju. Seetõttu kasutatakse vähegi suuremate valimimahtude korral ligikaudseid meetodeid olulisustõenäosuse leidmiseks. Üheks võimaluseks oleks Monte-Carlo kursusest tuttavate meetodite kasutamine. Teiseks võimaluseks oleks statistiku W asümptootilise jaotuse kasutamine.

2.2.1 Monte-Carlo meetod olulisustõenäosuse arvutamiseks

Meetodi idee on lihtne. Iga vaadeldud põllu puhul viskame juhuslikult kulli ja kirja ning otsustame takkajärgi, kumb põllupool peaks kandma silti "uus väetis". Saagikuste andmed jätame samaks — seega otsustab mündivise iga erinevuse ees kõigest tema märgi. Saagikuse ja väetise tüübi vahel seost olla ei tohi — oleme ju väetise tüüpi näitavad "sildid" juhuslikul segamini ajanud. Arvutame Wilcoxoni astakmärgitesti statistiku väärtuse sellise segaminiaetud andmestiku pealt. Kordame seda protseduuri palju kordi, näiteks miljon korda. Saame miljon statistiku W väärtust olukorras, kus nullhüpotees garanteeritult kehtib (tänu sellele, et vahetasime põllupooltel silte juhuslikult). Vaatame, kui sageli tuli meie poolt genereeritud andmestikes statistiku väärtus sama erakordne või veelgi erakordsem kui meie valimis. Saadud suhteline sagedus (sama veidrate või veelgi veidramate statistikute arv/ genereeritud valimite arv) olekski olulisustõenäosuse hinnanguks. Olulisustõenäosuse täpsemaks hindamiseks peaksime lihtsalt genereerima rohkem valimeid.

2.3 Asümptootiline jaotus

Astakmärgitesti statistik on kirja pandav kui juhuslike suuruste summa, $W = \sum_{i=1}^{n} Y_i$, kus $Y_i := \mathrm{sgn}(X_i)\mathrm{rank}(|X_i|)$. Juhul kui igal uuritaval objektil on valitud juhuslikult millises järjekorras me töötluseid rakendame (juhuslikult valime kumba haavadest saab plaastri; valime juhuslikult põllupoole kuhu külvame sorti A jne) saame liidetavaid lugeda sõltumatuteks. Sõltumatute juhuslike suuruste summa on üldjuhul asümptootiliselt normaaljaotusega (täpsemat tingimust vaatame kohe).

Eeldame, et nullhüpotees kehtib. Sellisel juhul tuleneb statistiku juhuslikkus puhtalt töötluste randomiseerimisest, seega suurused $\mathrm{rank}(|X_i|)$ on käsitletavad kui fikseeritud konstandid ja juhuslikud on ainult suurused $\mathrm{sgn}(X_i)$:

$$\begin{aligned} P(\mathrm{sgn}(X_i) = -1) &= 0.5 \\ P(\mathrm{sgn}(X_i) = 1) &= 0.5 \end{aligned}$$

Seega (nullhüpoteesi kehtides):

$$\begin{aligned} \mathrm{E}W &= \mathrm{E}\sum_{i=1}^{n} Y_i \\ &= \sum_{i=1}^{n} \mathrm{rank}(|X_i|)\mathrm{E}(\mathrm{sgn}(X_i)) \\ &= \sum_{i=1}^{n} \mathrm{rank}(|X_i|)0 \\ &= 0 \end{aligned}$$

ja

$$\begin{aligned} \mathrm{D}W &= \mathrm{D}\sum_{i=1}^{n} Y_i \\ &= \sum_{i=1}^{n} \mathrm{rank}(|X_i|)^2\mathrm{D}(\mathrm{sgn}(X_i)) \\ &= \sum_{i=1}^{n} \mathrm{rank}(|X_i|)^2 \\ &= 1^2 + 2^2 + 3^2 + \dots n^2 \\ &= n(n+1)(2n+1)/6 \end{aligned}$$

Vahemärkus: Võrduse $\sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/6$ kehtivuses veendumiseks võib kasutada näiteks induktsiooni. Veendu, et antud valem kehtib, kui $n = 1$. Siis näita, et kui valem kehtib n korral, siis kehtib ta ka $n + 1$

korral:

$$
\begin{aligned}
\sum_{i=1}^{n} i^2 &= \sum_{i=1}^{n} i^2 + (n+1)^2 \\
&= n(n+1)(2n+1)/6 + (n+1)^2 \\
&= (n+1)(2n^2+n)/6 + (n+1)^2 \\
&= (n+1)(2n^2+n+6n+6)/6 \\
&= (n+1)(2n^2+7n+6)/6 \\
&= (n+1)(n+2)\left(2(n+1)+1\right)/6
\end{aligned}
$$

Üritame nüüd veenduda, et suurte valimite (suur n) korral on statistiku W jaotus ligilähedaselt normaaljaotusega. Tõestusel tugineme Lindeberg-Feller'i teoreemile.

Olgu antud sõltumatute juhuslike suuruste seeriad (igas reas olevad juhuslikud suurused olgu omavahel sõltumatud, veerud võivad olla sõltuvad või osaliselt isegi kokku langeda):

$$
\begin{aligned}
&X_{11} \\
&X_{21}, X_{22} \\
&X_{31}, X_{32}, X_{33} \\
&\dots
\end{aligned}
$$

Olgu $\mathrm{E}X_{ij} = 0$, $\mathrm{D}X_{ij} = \sigma_{ij}^2$, $S_n := \sum_{j=1}^{n} X_{nj}$, $\mathrm{D}S_n := \sum_{j=1}^{n} \sigma_{nj}^2$.

Lindeberg-Felleri teoreem väidab: Kui mistahes $\varepsilon > 0$ korral

$$
\frac{1}{\mathrm{D}S_n} \sum_{j=1}^{n} \mathrm{E}(X_{nj}^2 I(|X_{nj}| \geq \varepsilon \mathrm{D}S_n)) \to 0, \text{ kui } n \to \infty
$$

siis

$$
S_n/\sqrt{\mathrm{D}S_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).
$$

Wilcoxoni astakmärgitesti puhul oleksid X_{ni} rollis märgiga astakud. Valimi suuruse n korral oleks $X_{ni} = i$ või $X_{ni} = -i$:

valim suurusega 1	± 1
valim suurusega 2	$\pm 1, \pm 2$
valim suurusega 3	$\pm 1, \pm 2, \pm 3$
	$\dots$

Samas reas olevad juhuslikud suurused on nullhüpoteesi kehtides sõltumatud, sest miinus või plusmärgi sattumise astaku ette otsustab randomiseerimine (ja loodetavasti on kõik põldude randomiseerimiseks toimunud mündivisked teineteisest sõltumatud...).

Valimi suurusega n puhul oleks Wilcoxoni astakmärgitesti statistiku väärtus seega $W_n = \sum_{i=1}^n X_{ni}$.

Kuna liidetavad Wilcoxoni testi puhul on (nullhüpoteesi kehtides) juhuslikud ainult märgi poolest, on $|X_{nj}|$ ja X_{nj}^2 tegelikult konstandid (astak ja astaku ruut). Kuna aga maksimaalne astak on n, siis $|X_{nj}| \leq \max_{j\leq n} |X_{nj}| = n$ ja järelikult:

$$
\begin{aligned}
\frac{1}{\mathrm{D}S_n}\sum_{j=1}^n \mathrm{E}(X_{nj}^2 I(|X_{nj}| \geq \varepsilon \mathrm{D}S_n)) &\leq \frac{1}{\mathrm{D}S_n}\sum_{j=1}^n \mathrm{E}(X_{nj}^2 I(n \geq \varepsilon \mathrm{D}S_n)) \\
&= I(n \geq \varepsilon \mathrm{D}S_n)\frac{1}{\mathrm{D}S_n}\sum_{i=1}^n i^2 \\
&= I(n \geq \varepsilon \mathrm{D}S_n) \\
&= I(n/\mathrm{D}S_n \geq \varepsilon)
\end{aligned}
$$

Viimane avaldis koondub aga nulliks:

$$
n/\mathrm{D}S_n = \frac{n}{n(n+1)(2n+1)/6} \to 0 \text{ kui } n \to \infty.
$$

Järelikult on Lindebergi tingimus rahuldatud ja nullhüpoteesi kehtides $W/\sqrt{\mathrm{D}W} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

Näide 2.2 *Proovitakse, kas päikesekreem ikka kaitseb nahka päikese eest. Katsealustel määritakse lapike nahka kreemiga kokku. Saadetakse nad siis randa. Hiljem mõõdetakse naha punetust nii kreemitatud kohast kui ka selle kõrvalt. Naha punetuse erinevused (kaitsmata koht-kreemitatud koht) on järgmised:*

-1, -3, 4, -6, 7, 8, -9, -10, 12, 14, 15, -17, 20, 25, 29, 30, -32, 40, 78, 102.

Wilcoxoni astakmärgitesti statistiku väärtuseks saame:

$$
\begin{aligned}
W &= -1-2+3-4+5+6-7-8+9+10+ \\
&\quad +11-12+13+14+15+16-17+18+19+20 = 108
\end{aligned}
$$

Nullhüpoteesi kehtides ootaksime antud valimi suuruse juures statistiku keskväärtuseks ja dispersiooniks $EW = 0$ *ja* $DW = n(n+1)(2n+1)/6 = 20 * 21 * 41/6 = 2870$*. Seega peaks nullhüpoteesi kehtides juhuslik suurus* $(W - EW)/\sqrt{DW} = W/53,57$ *olema ligikaudu standardse normaaljaotusega juhuslik suurus (nullhüpoteesi kehtides peaks vastav suurus ootuspäraselt jääma vahemikku -1,96...1,96). Antud valimi korral aga* $W/53,57 = 108/53,57 = 2,016$*, seega midagi tundub olevat viltu... nullhüpoteesi eeldusega. Olulisustõenäosuse saame, kui leiame, kui tõenäoliselt oleksime näinud sedavõrd ekstreemset või veel ekstreemsemat statistiku väärtust nullhüpoteesi kehtides. Tõenäosus näha veel suuremat statistiku väärtust* H_0 *kehtides oleks* $1 - \Phi(2,016) = 0,0219$ *(olulisustõenäosus ühepoolse hüpoteesi korral). Kahepoolse hüpoteesi korral peaksime muidugi siia juurde liitma ka erakordselt väikese* W *saamise tõenäosuse (nullhüpoteesi kehtides). Normaaljaotuse sümmmetriat arvestades saame tulemuseks* $(1 - \Phi(2,016)) * 2 = 0,0438$*. Seega on olulisustõenäosus väike ja me võime nullhüpoteesi (olulisuse nivool 0,05) kummutatuks lugeda. Võrdlusena: täpne olulisustõenäosus oleks 0,044054.*

2.3.1 Pidevuse parandus

Statistikas tuleb sageli ette olukordi, kus me lähendame diskreetset jaotust pideva jaotusega. Näiteks võime lähendada binoomjaotust normaaljaotusega või Wilcoxoni astakmärgitesti statistiku jaotust normaaljaotusega. Kui meie valimid on väga suured, siis võime sellist lähendamist teha üsna sirgjooneliselt ja probleemideta. Keskmise suurusega valimite korral saame aga sageli veidi täpsemaid tulemusi, kui enne arvutuste alustamist veidi mõtleme.

Vaatame näiteks juhuslikku suurust $X \sim B(10; 0,4)$. Antud binoomjaotust saab lähendada normaaljaotusega juhusliku suurusega $Y \sim N(4, \sigma^2 = 2,4)$. Kuidas leida normaaljaotuse lähendit kasutades tõenäosust, et $X = 1$, $P(X = 1) = 0,0403$? On selge, et antud tõenäosus pole ligilähedaseltki võrdne tõenäosusega $P(Y = 1) = 0$. Parem variant antud tõenäosuse leidmiseks oleks vaadata tõenäosust $P(0.5 \leq Y \leq 1.5) = 0,0414$.

Kuidas aga oleks mõistlik leida normaaljaotuse lähendit kasutades suurust $P(X \leq 1) \approx 0,04636$? Me võime mõelda ka nii: $P(X \leq 1) = P(X = 1) + P(X = 0)$. Viimast tõenäosust aga lähendab tõenäosus $P(Y \leq 1.5) = 0,0533$ palju täpsemalt kui tõenäosus $P(Y \leq 1) = 0,026$. Seega soovitatakse kasutada diskreetsete juhuslike suuruste lähendamisel pidevuse korrektsiooni (*continuity correction*): $P(X \leq c) \approx P(Y + 0,5 \leq c)$ või $P(X \geq c) \approx P(Y - 0,5 \geq c)$.

Vaatame pidevuse parandust eelnenud näite varal. Leidmaks tõenäosust $P(W \geq 108)$ oleks meil seega soovitav vaadata normaaljaotuse jaotusfunktsiooni kohal $(W-1)/53,57$ (kuna W muutub meil sammuga 2), mis annaks kahepoolse hüpoteesi puhul olulisustõenäosuseks $(1-\Phi(1,9974))*2 = 0,0458$. Antud näites ei andnud pidevuse paranduse kasutamine täpsemat tulemust — kui ilma pidevuse paranduseta alahindasime olulisustõenäosust, siis pidevuse parandust kasutades saime liiga suure hinnangu olulisustõenäosusele. Lähend on kõigest lähend ja vahel võib pidevuse paranduse kasutamine ka halvemaid tulemusi anda. Enamasti saadakse siiski pidevuse parandust kasutades täpsemaid tulemusi (ja ka antud näite puhul on arvatavasti konservatiivsem hinnang olulisustõenäosusele eelistatum).