

3.6 Wilcoxon astaksummatesti tulemuse tõlgendamisest

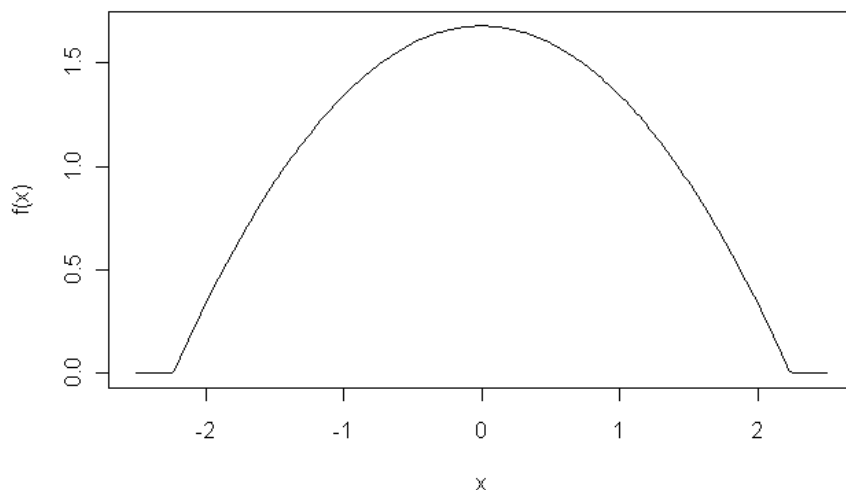
Kui saame Wilcoxon astaksummatesti abil lugeda tõestatuks alternatiivse hüpoteesi, siis mida see ikkagi tähendab? Kirjanduses eksisteerib väga palju eksitavaid väiteid selle kohta, mida Wilcoxon astaksummatest siis ikkagi kontrollib. Nii nagu leidub eksitavaid väiteid ka selle kohta, millised on Wilcoxon astaksummatesti eeldused. Näiteks kohtab väiteid, mis väidavad, et Wilcoxon astaksummatest kontrollib keskväärtuste või mediaanide erinevust.

See väide ei pea paika.

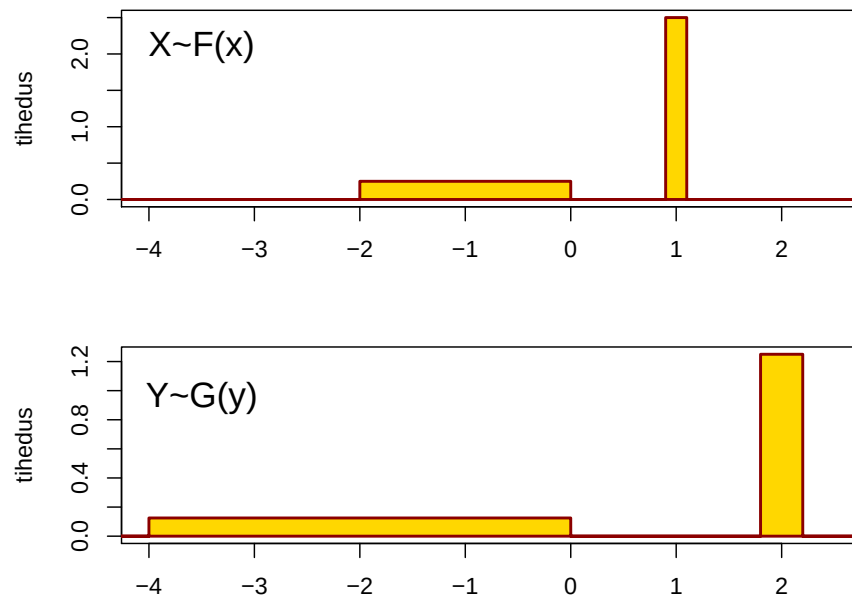
Saab tuua näite olukorrast, kus mõlemas kontrollitavas populatsioonis on nii uuritava tunnuse keskväärtused kui ka mediaanid võrdsed, kuid Wilcoxon astaksummatest võtab kindlalt vastu alternatiivse hüpoteesi (vähegi mõistlikuma valimisuuruse juures). Vaata näiteks joonist 3.3.

Asjakohasem oleks ikkagi öelda: kui võtame ühest populatsioonist ühe juhuslikult valitud esindaja X ja teisest populatsioonist ühe juhuslikult valitud esindaja Y , siis Wilcoxon astaksummatest kummutab nullhüpoteesi

Joonis 3.2: Jaotus, mille korral $e_{t,W} = 0,864$ (maksimaalselt t-testi soosiv jaotus).



Joonis 3.3: Kaks jaotust, mille võrdlemisel Wilcoxon astaksummatest võtab vastu alternatiivse hüpoteesi, kuigi mõlema jaotuse mediaan ja keskvärtus on samad.



siis, kui $p_1 := P(X < Y) \neq 0,5$. Ülaltoodud näite korral on $p_1 \approx 0,56$, seega on ka alternatiivse hüpoteesi vastuvõtmine Wilcoxon testi poolt õigustatud. Vahel mainitaksegi, et hüpotees mida Wilcoxon astaksummatest kontrollib, on väide $H_0 : P(X < Y) = 0,5$.

Samuti võib segadustekitavalt mõjuda mõnede allikate väide, et Wilcoxon testi tohib kasutada vaid siis, kui $F(x) = G(x + \Delta)$ ehk nn nihkealternatiivi korral. Kui tegemist on jaotuse nihkega, siis saab Wilcoxon astaksummatesti tõepoolest interpreteerida kui testi mediaanide (või keskvärtuste, kui keskvärtus eksisteerib) võrdlemiseks. Paraku on väga raske kindlaks teha, kas vastav tingimus on täidetud. Samuti on ta suuresti mittevajalik — kui antud tingimus pole täidetud, ei juhtu Wilcoxon testiga midagi muud, kui vaid see, et me ei tohi enam antud testi interpreteerida kui testi mediaanide (või keskvärtuste) võrdlemiseks.

Kui Wilcoxon astaksummatest võtab vastu alternatiivse hüpoteesi, siis efekti suurust saab iseloomustada kasutades hinnangut suurusele $P(X < Y)$, sobivaks hinnanguks oli $W_{xy}/(m \cdot n)$.

3.7 Võrdsed vaatlused

Kui leidub vaatluseid, mida pole võimalik järjestada, kasutatakse taas keskmiseid astakuid ehk *midrank*'e. Loomulikult muutub seetõttu ka Wilcoxon'i astaksummatesti teststatistiku jaotus (näiteks võib tekkida võimalus näha mittetäisarvulist teststatistiku väärtust). Kui leiame teststatistiku jaotust kõigi võimaluste läbivaatamise teel, pole sellest suuremat häda (tuleb vaid mees pidada, et näiteks keskmiste astakute 3,5 ja 3,5 puhul on tegemist kahe erineva numbriga, seega astakute kombinatsioonid $\{1; 3, 5_1; 7\}$ ja $\{1; 3, 5_2; 7\}$ on kaks erinevat kombinatsiooni).

Kui soovime kasutada teststatistiku asümptootilist jaotust, siis tuleb mees pidada, et teststatistiku dispersioon muutub tänu keskmiste astakute kasutamisele:

$$DW_{xy} = DW_s = \frac{mn(N+1)}{12} - \frac{mn \sum_i^e (d_i^3 - d_i)}{12N(N-1)},$$

kus e on erinevate väärtuste arv ühendatud valimis ja d_i on i . unikaalse väärtuse korduste arv (mitu korda seda väärtust esines kahe võrreldava valimi peale kokku).

Märkus: statistiku W_s (või W_{xy}) jaotus ei pruugi olla normaaljaotusele lähedane isegi suure valimi korral, kui $\max_i d_i/N \approx 1$ ehk kui mõni konkreetne uuritava tunnuse väärtus domineerib mõlemas valimis.

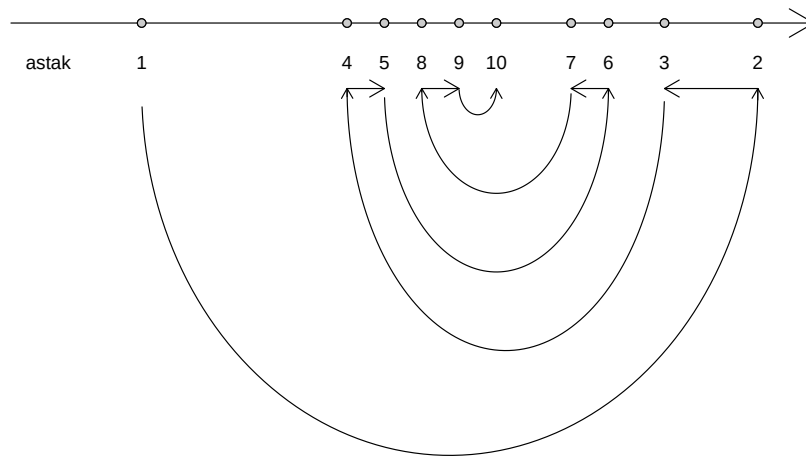
Märkus 2: pidevuse paranduse kasutamine on keeruline, kui valimites esineb samasuuri uuritava tunnuse väärtuseid — teststatistiku võimalike väärtuste vahelised kaugused hakkavad sellises olukorras varieeruma. Seega +0,5 ja −0,5 asemel võib tarvis minna ka +0,25 või −0,25 või Ühesõnaga, võrdse suurusega vaatluseid sisaldava andmestiku korral pidevuse parandust enamasti (praktilistel põhjustel) ei kasutata.

3.8 Siegel-Tukey ja Ansari-Bradley testid

Wilcoxon'i testi puhul on alternatiivne hüpotees $F \neq G$ väga üldine. Osade konkreetsete alternatiivide korral on Wilcoxon'i testi võimsus madal. Näiteks kuigi $F \neq G$, aga kui $P(X < Y) \approx 0,5$, siis Wilcoxon'i testi võime alternatiivse hüpoteesi kehtivust tõestada on väga madal. Selline situatsioon võib näiteks tekkida siis, kui kaks võrreldavat populatsiooni erinevad eelkõige selle poolest, et ühes populatsioonis on uuritava tunnuse hajuvus suurem kui teises (suur geneetiline varieeruvus metsikus populatsioonis vs loomaaialoode või aretatud taimesortide ühetaolisus; rikkis kontrollaparatuur ei suuda

tagada toodangu ühetaolisust jne). Parandamaks testi võimet eristada just nimelt erinevust varieeruvuses, võime muuta astakute paneku järjekorda. Väikseim väärtus saab astaku 1, suurim väärtustest saab astaku 2, suuruselt teine vaatlus saab astaku 3 jne, vaata ka joonist 3.4. Teststatistikuks on sarnaselt Wilcoxononi astaksummatestile ühe valimi vaatluste astakute summa. Saadud statistiku jaotus nullhüpoteesi kehtides on sama mis Wilcoxononi testil.

Joonis 3.4: Astakute jaotamine Siegel-Tukey testi korral



Siegel-Tukey testile heidetakse ette, et testi statistiku väärtus (ja testi tulemus) võib sõltuda sellest, kummast otsast (suurimast või väikseimast väärtusest) astakute jagamist alustatakse. Variatsioonrea mõlemaid otsi kohtleb poliitilisemalt korrektsemalt (võrdsemalt) Ansari-Bradley test. Nimetatud testi puhul antakse nii suurimale kui ka väiksemale väärtusele astak 1, variatsioonrea 2. ja eelviimasele vaatlusele astak 2 jne. Taolisel viisil saadud astakute summa ühe grupi jaoks annabki Ansari-Bradley teststatistiku väärtuse. Kahjuks pole aga Ansari-Bradley teststatistiku jaotus nullhüpoteesi kehtides aga sama kui Wilcoxononi testil.

Märkus: Siegel-Tukey ja Ansari-Bradley testide võime tuvastada hajuvuse erinevust kahaneb, kui lisaks erinevatele hajuvustele on ühes populatsioonis uuritava tunnuse väärtused suuremad kui teises. Üks praktikas kasutatav lahendus nimetatud probleemile oleks järgmine: hinnatakse Wilcoxononi teststatistikut kasutades võimalikku nihet Δ , mille võrra tuleks populatsioone

nihutada, et $P(X < Y) = 0.5$. Tehaksegi soovitud nihe, $Y_{uus} = Y - \Delta$. pärast tehakse siis kas Siegel-Tukey või Ansari-Bradley test nihutatud väärtuseid kasutades. Nimetatud testprotseduur pole päris korrektne, aga suure valimi puhul on võimalikud probleemid on tühised ja ignoreeritavad.

Näide 3.3 *Kahtlustatakse, et teatud ainega X kokkupuutel kaob inimorganismi võime kontrollida β -glükoosi taset veres. Aga β -glükoosi taseme järsud muutused võivad kahjustada maksa ja põhjustada ajapikku tõsiseid probleeme. Otsustatakse testida, kas ainega X kokkupuutunud inimestel on β -glükoosi kontsentratsioon varieeruvam kui kontrollgrupi inimestel.*

<i>Eksponeeritud</i>	67, 80, 120
<i>Kontrollgrupp</i>	78, 82, 84, 86, 90, 92, 95

Siegel-Tukey teststatistik on $W_{ST} = 1 + 2 + 5 = 8$. Veel väiksema teststatistiku väärtuse saamise võimalusi on 4: $\{1, 2, 3\}$; $\{1, 2, 4\}$; $\{1, 2, 5\}$; $\{1, 3, 4\}$. Kokku on võimalusi näpsata 3-e astakut 10 seast $C_{10}^3 = 120$. Ühepoolse testi olulisustõenäosus on seega $4/120 = 0,0333...$ ja kahepoolse testi olulisustõenäosus on $0,0666...$

Ansari-Bradley testi korral oleks teststatistiku väärtuseks $W_{AB} = 1 + 1 + 3 = 5$, astakute komplekte, mis annaksid samasuure või väiksema statistiku väärtuse, oleks 6 (kuna iga astakut esineb kaks korda, siis eristame neid alaindeksitega v ja p): $\{1_v, 1_p, 2_v\}$; $\{1_v, 1_p, 2_p\}$; $\{1_v, 1_p, 3_v\}$; $\{1_v, 1_p, 3_p\}$; $\{1_v, 2_v, 2_p\}$; $\{1_p, 2_v, 2_p\}$. Ühepoolse testi olulisustõenäosus tuleks seega $6/120 = 0,05$ ja kahepoolse testi korral $0,1$.

Ülesanded

1. Kõige raskemate, viimases staadiumis vähihaigete peal katsetati eksperimentaalset ravimeetodit. Osad patsientidest randomiseeriti eksperimentaalset ravi saavasse patsientide gruppi, osad tavapärasest ravi saavasse kontrollgruppi. Patsientide elulemused (kuudes) mõlemas grupis olid järgmised:

Eksperimentaalne grupp:

28.6 11.4 7.9 10.7 6.1 5.6 10.8 44.6 16.7 15.0 19.4 5.4 8.2 18.2 7.0 15.2
8.0 12.3 12.5 7.4 15.8 15.3 17.9 17.5 10.5

Kontrollgrupp:

7.6 11.8 1.5 1.4 4.4 28.9 12.3 5.4 9.6 1.5 13.9 7.6 12.4 44.2 10.5 10.4

18.8 6.5 3.4 5.9

Leia Wilcoxon'i astaksummatesti abil, kas elukestvuse jaotus on erinev nendes kahes patsientide grupis. Kui näed erinevust, siis milline see erinevus on (kumb ravimeetod töötas paremini?). Võrdle ravimeetodeid ka t-testi abil, kommenteeri saadud tulemust (mida saad t-testi tulemuseks ja miks).

2. Milline on Ansari-Bradley teststatistiku jaotus H_0 kehtides, kui $m = 3$ ja $n = 4$?
3. Soovitakse testida, kas hommikupudru söömine aitab lapsel pikaks kasvada. Lastekodus elavad väikelapsed kavatakse juhuslikult jagada kahte rühma. Üks rühmadest saaks hommikusöögiks putru, teine rühmadest muud hommikusööki (hommikuhelbeid, võileibu vms). Kui lapsed saavad ükskord 16 aasta vanusteks, siis mõõdetakse kõigi laste pikkused. Kokku on sobivas eas lapsi 50 (seega saaks nad jagada kahte 25-lapselisse rühma). Kavandatav uuring on pikaajaline ja vaevarikas. Sellise uuringu õigustamiseks tuleb eelnevalt teada, kas kogutavate andmete põhjal on lootust midagi põhjapanevat ja veenvat öelda hommikupudru mõju kohta lapse pikkusele. Kuna uuringuandmete analüüsiks on kavas kasutada Wilcoxon'i astaksummatesti, tuleks seega leida astaksummatesti võimsus. Arvuta Wilcoxon'i astaksummatesti võimsus antud uuringu puhul!

Märkus: Selle ülesande lahendamiseks pead tegema teatud valikuid ja kasutama lisainformatsiooni uuritava tunnuse ja töötluse (pudrusöömine) kohta. Inimese pikkus ja toitumine on valdkonnad, mille kohta kõik inimesed paistavad midagi teadvat — seega saad sa ise võimaluse olla eksperdi rollis ja välja pakkuda oma eksperthinnangud suurustele, mida on võimsuse arvutamiseks tarvis teada. Soovitavalt muidugi koos põhjendusega - vajadusel viidetega kirjandusele/internetiallikele jne. Päris absurdseid väärtuseid ei tasu ka pakkuda — kõiki valikuid tuleb ju tegeliku projekti korral projekti rahastajate ees ka kaitsta!

4. (vabatahtlik) Wilcoxon'i astaksummatesti teststatistiku arvutamisel võime astakud asendada ka millegi muuga, näiteks võime suuruselt i . vaatluse puhul astaku i asendada mingi muu i -st sõltuva arvuga, $i \rightarrow a_i$ (näiteks i^2 -ga). Näiteks kasutatakse sageli (van der Waerden'i test) astakute asemel suuruseid $a_i = \Phi(i/(N+1))$, aga võime ka eelistada mõnda muud valikut. Ka teiste a_i valikute puhul on teststatistiku $W_a = \sum_{i \in \text{Treatment}} a_i$ asümptootiliseks jaotuseks nullhüpoteesi

kehtides (enamasti) normaaljaotus,

$$\frac{W_a - EW_a}{\sqrt{DW_a}} \sim N(0, 1).$$

Paraku on asümptootilise jaotuse teadmisest kasu vaid siis, kui oskame leida EW_a ja DW_a väärtuseid nullhüpoteesi kehtides. Kuidas avalduvad mainitud suurused suuruste $\bar{a} := 1/N \sum_{i=1}^N a_i$ ja $s_a^2 := 1/N \sum_{i=1}^N (a_i - \bar{a})^2$ kaudu?