Wilcoxons Mann-Whitney prófið (*Wilcoxon Rank-Sum test*)

Fyrirlestur í Tölfræði II (SÁL203G)

2009-02-02

Hvenær er *t*-próf ótraust?

t-próf hefur reynst traust (*robust*) gagnvart höfnunarmistökum. Ef hóparnir eru nægjanlega stórir og jafnir, fæst rétt niðurstaða jafnvel þótt dreifing mæligilda sé umtalsvert skekkt.

t-próf er ótraust gagnvart frávillingum sérstaklega í litlum hópum. Stakt jaðargildi í stórum hópi þarf þó ekki að koma að sök.

t-próf er einnig ótraust ef dreifingar eru ólíkar, þ.e. jákvætt skekkt í öðrum hópnum en neikvætt í hinum.

Þótt *t*-próf sé traust og afkastamesta prófið á mismun meðaltala *þegar* forsendur standast, þá geta aðrar aðferðir verið mun afkastameiri ef frávik eru frá normaldreifingu eru veruleg.

Raðsummupróf Wilcoxons—eða Wilcoxon Mann-Whitney—er nánast jafnnæmt á frávik ef forsendur *t*-prófs standast en oft mun afkastameira ef þær standast ekki.

© 2005–2009 Guðmundur B. Arnkelsson

Viðbrögð við brotum á forsendum

Stök jaðargildi má fjarlægja ef það er bersýnilegt að þau eru tilkomin með öðrum hætti en önnur mæligildi. Þetta geta t.d. verið skráningarmistök eða stök úr öðru þýði, t.d. 15 ára piltur á námskeiði um ættfræði.

Ef ekki er hægt að endurskilgreina frávilling gæti komið til greina að fjarlægja hann samt. Þá fer best að lýsa áhrifum hans á niðurstöðuna.

Almennt er gagnrýnisvert að fjarlægja mæligildi án skýrs rökstuðnings.

Umbreyting lagar oft skekkju. Stundum hverfa frávillingar eða verða minna ýktir við slíka umbreytingu.

Nýjar úrvinnsluaðferðir veita stundum úrlausn. Þannig er hægt að gera ráð fyrir annarri dreifingu heldur en normaldreifingu eða meta marktekt og öryggisbil með endurúrtökum (*bootstrapping*).

Að síðustu má nota úrtakshháð (stikalaust; *nonparametric*) próf af ýmsu tagi.

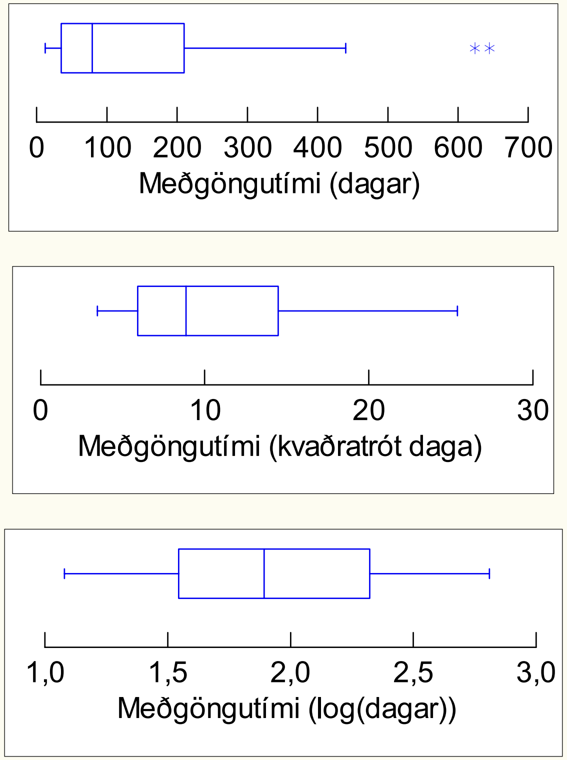
© 2005–2009 Guðmundur B. Arnkelsson

Hvað eru umbreytingar?

Umbreytingar eru það þegar við notum ekki upprunalegar einingar gagnanna heldur breytum þeim í nýjar.

Myndirnar sýna meðgöngutíma ólíkra dýrategunda.

Ef hann er í dögum, fáum við tvö áberandi jaðargildi og jákvætt skekkta dreifingu. Með kvaðratrót minnkar skekkja og jaðargildi hverfa. Með lógariþma fæst nánast samhverf dreifingu án jaðargilda.

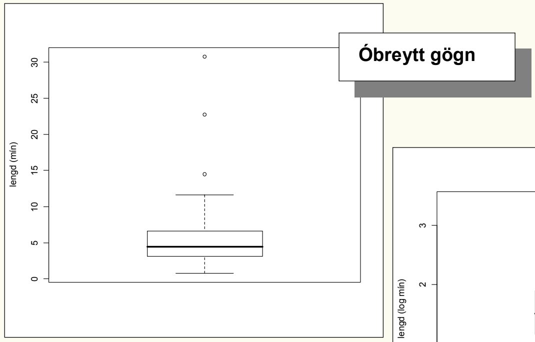


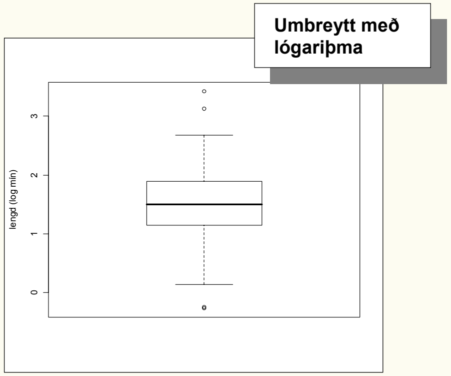
Lengd laga í Ipod McCabe

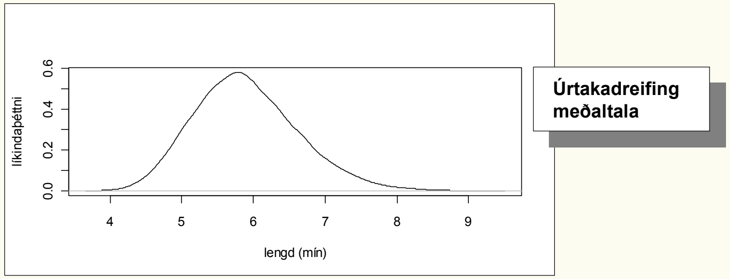
Myndirnar sýna 50 laga úrtak úr Ipod McCabe. Dreifingin er lítillega skekkt og með þremur áberandi frávillingum.

Lógariþmaumbreyting fjarlægir skekkju og dregur úr frávillingum. Þannig henta gögnin betur til úrvinnslu sbr. skekkta úrtakadreifingu óbreyttra gagna.

Umbreytt 95% öryggisbil er 1,3–1,7 log mínútur sem er á illskiljanlegum kvarða, samanborið við 4,4–7,4 mínútur fyrir óbreytt gögn.







© 2005–2009 Guðmundur B. Arnkelsson

Raðsummupróf Wilcoxons

Raðsummupróf Wilcoxons byggir á þeirri hugmynd að ef hóparnir eru eins ættu raðtölurnar að vera svipaðar.

Taflan sýnir uppskeru eftir því hve mikill arfi er í beðum. Ef hóparnir eru eins, ætti tilviljun að ráða því hvar í röðinni mæligildin lenda. Taflan sýnir hins vegar að mælingar þar sem 3 hélunjólar eru á hvern metra beðsins fá að jafnaði lægri raðtölu heldur en þegar allt illgresi hefur verið reytt.

Samræmist þetta því að hélunjóli hafi ekki áhrif á uppskeru maís.

Arfi	Uppskera	Röð
3	153,1	1
3	156,0	2
3	158,6	3
0	*165,0*	*4*
0	*166,7*	*5*
0	*172,2*	*6*
3	176,4	7
0	*176,9*	*8*

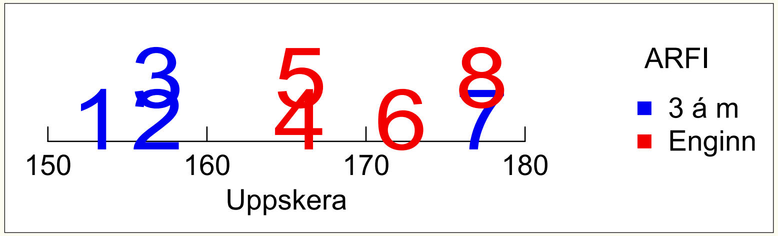
Raðtölur eru myndaðar með því að láta lægsta mæligildið fá 1, það næsta 2, o.s.frv. Ef tvö eða fleiri mæligildi eru eins, er notað meðaltal viðkomandi raðtalna.

© 2005–2009 Guðmundur B. Arnkelsson

Hvaða núlltilgáta er prófuð?

Ef dreifingin er eins í báðum hópum, má líta svo á að raðsummupróf prófi þá núlltilgátu að miðtölurnar séu þær sömu.

Almennt séð er þó prófað hvort dreifingarnar séu eins. Ef núlltilgátan er röng, kemur annar hópurinn úr þýði með mæligildi sem að jafnaði eru hærri. Það getur verið að það sé staðsetning dreifingarinnar sem er ólík (ólíkar miðtölur) en það geta einnig verið aðrir eiginleikar sem gera mæligildin hærri.



Prófið mælir hvort raðtölur séu kerfisbundið hærri eða lægri í öðrum hópnum. Þar sem þetta eru raðtölur, dregur mjög úr áhrifum frávillinga. Kröfur til dreifingar (svo sem normaldreifing) hverfa einnig, þar sem raðtölur hafa alltaf sömu dreifingu, t.d. hér tölurnar 1, 2, 3, …8.

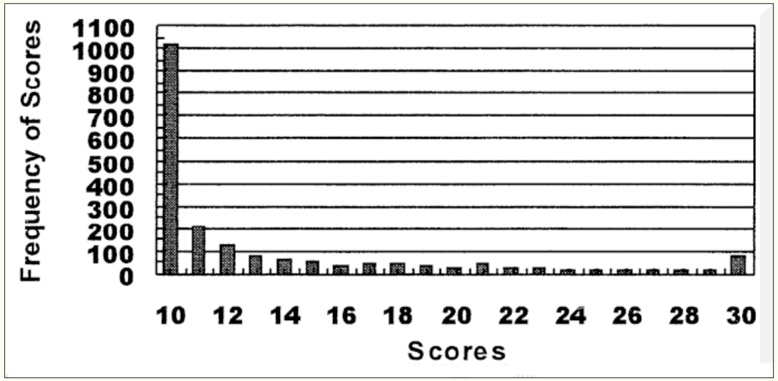
© 2005–2009 Guðmundur B. Arnkelsson

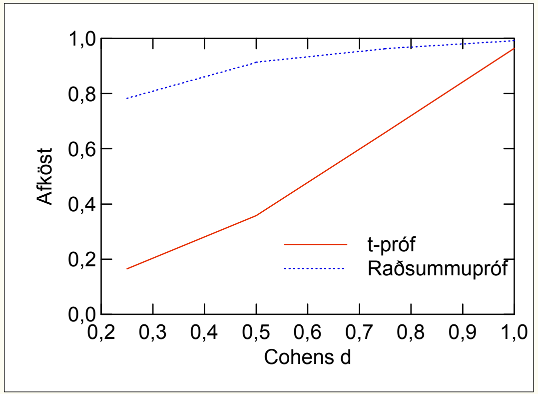
Kostir Wilcoxon Mann-Whitney

Wilcoxon Mann-Whitney gerir minni kröfur til dreifingar mæligilda heldur en *t*-próf. Þetta kemur sér vel ef áberandi frávillingar eru í gögnunum eða dreifing mæligilda víkur áberandi frá normaldreifingu.

Prófið er með svipuð afköst og *t*-próf við normaldreifingu en getur verið mun afkastameira ef skekkja er mikil eða frávillingar.

Ef dreifing er eins í báðum hópum, er munur á staðsetningu hópanna prófaður svipað og í *t*-prófi.





Bridge, P.D., & Sawilowsky, S.S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: Comparative power of the t-test and wilcoxon rank-sum test in small samples Applied Research. *Journal of Clinical Epidemioly, 52*, 229–235.

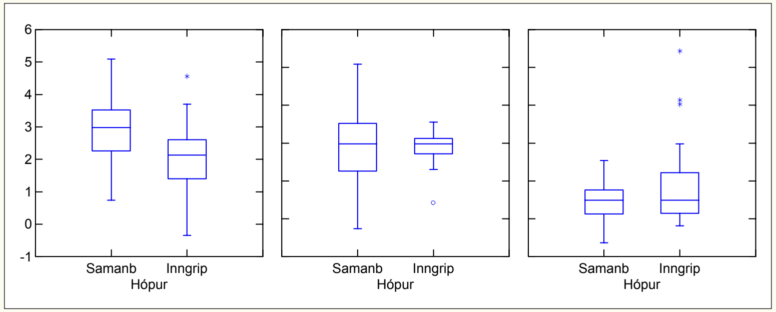
© 2005–2009 Guðmundur B. Arnkelsson

Gallar Wilcoxon Mann-Whitney

Wilcoxon Mann-Whitney prófar aðeins mun á staðsetningu ef dreifing er eins í báðum hópum.

Almennt séð prófað það hvort mæligildi annars hópsins hneigist til að vera hærri en mæligildi hins hópsins. Það getur gerst jafnvel þó t.d. miðtölur séu eins.

Ef breidd (*spread*) dreifinganna er ólík undir núlltilgátunni, verður Wilcoxon Mann-Whitney ótraust. Prófið er því næmt fyrir breyttri staðsetningu en einnig breidd.



Fyrsta myndin sýnir ólíka staðsetningu og því er H_0 röng. Á miðmyndinni er sama staðsetning en mjög ólík breidd. Hér gæti Wilcoxon Mann-Whitney verið ótraust og hafnað ranglega oftar en α. Á síðustu mynd er mismunandi lag á dreifingu í hópunum (og ólík breidd líka). Hér er núlltilgátan röng þrátt fyrir sömu staðsetningu.

© 2005–2009 Guðmundur B. Arnkelsson

Útreikningur

Við breytum mæligildum í raðtölur þannig að lægsta talan fái 1, næsta fái 2 o.s.frv.

Við leggjum saman raðtölurnar í hvorum hópi fyrir sig.

Við getum flett upp ýmist lægri eða hærri summunni í viðeigandi töflu, passa að gera það rétt! Þá er niðurstaðan örugglega rétt. Við getum líka notað normalnálgunina sem sýnd er hér til hliðar, gjarnan með leiðréttingu fyrir (skort á) samfellu.

W : Summa raðtalna í hópi 1

$$\mathrm{N} = n_1 + n_2$$

$$\mu_W = \frac{n_1(N+1)}{2}$$

$$\sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}$$

$$z = \frac{W - \mu_W}{\sigma_W}$$

Samfelluleiðrétting (continuity correction)

$$z = \frac{W - 0{,}5 - \mu_W}{\sigma_W}$$

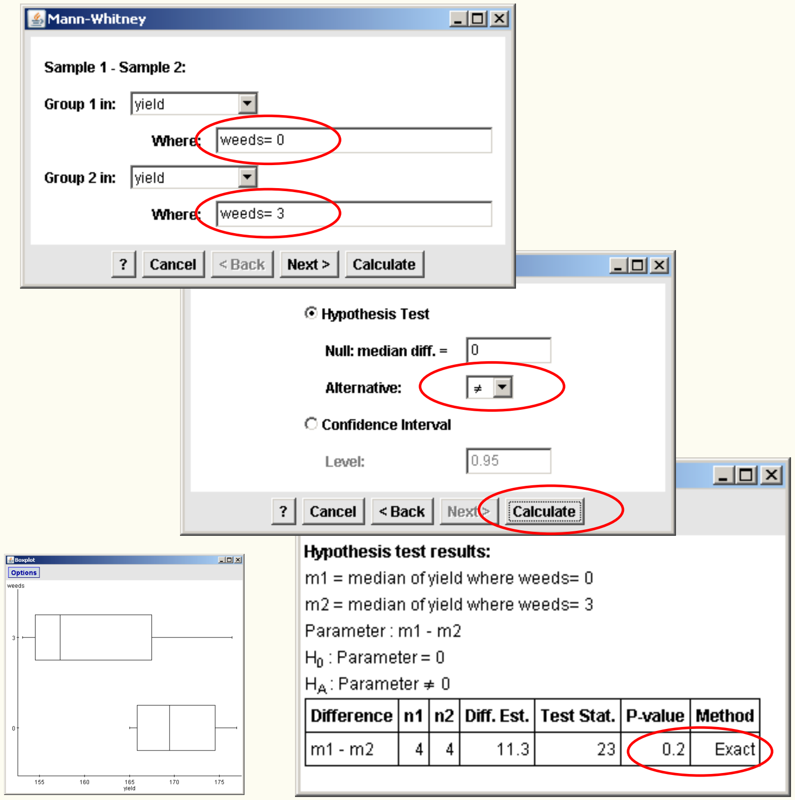
© 2005–2009 Guðmundur B. Arnkelsson

WMW-próf í CrunchIt

Ég les `eg15_001.txt` inn. WMW-próf er framkvæmt með því að fara í `Stat/ Nonparametrics/ Mann-Whitney` Þar set ég `yield` fyrir báða hópanna en skilgreini þá hvorn fyrir sig með `Where` og smelli á `Calculate`.

Hægt er að biðja um öryggisbil en óljóst hve gagnlegt það er.

Forsendur má meta með `Graphics/ Boxplot`, `yield` sem fylgibreytu og `weeds` í `Group by`. Haka þarf við `Use fences to identify outliers`. Rétt er að vera strangur varðandi misleitni.

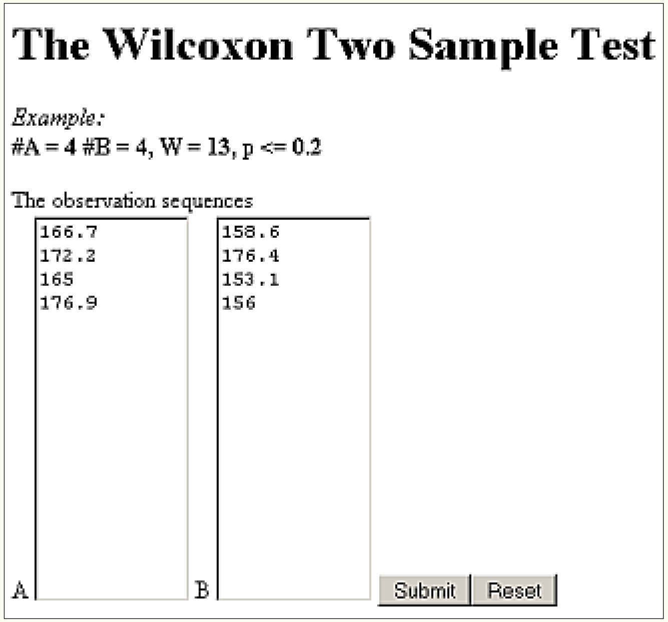


Vefreiknir fyrir Wilcoxon

Með vefreikninum er hægt að setja inn hrátölurnar, hvorn hóp í sinn dálkinn. Síðan er smellt á `Submit` og þá birtast niðurstöður fyrir ofan dálkanna.

Þessi vefreiknir notar hárrétta (*exact*) útreikninga þegar hóparnir eru litlir en skiptir yfir í normalnálgun þegar þeir stækka.

Marktektargildið miðast við að prófið sé tvíhliða.



www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html

© 2005–2009 Guðmundur B. Arnkelsson