

Ikke-parametriske metoder

12.11.2008 Eirik Skogvoll, AKF

Rosners kap. 9 om ikke-parametriske metoder er uoversiktlig, og det er nesten utelukkende anvendt normaltilnærming med relativt stor utvalgsstørrelse. Dette kan virke litt forvirrende, idet det er ved "liten n " at metodene er lettest å forstå.

Nedenfor er et utdrag fra Rosners tabell 8.18, s. 342, relatert til Problem 9.7, s. 380. Tabellen angir fødselsvekt (wt , målt i pund; 1 lb = 454 g) til barn hvis mødre har fått et medikament som sies å forhindre prematur fødsel (trt), eller placebo (ctr).

trt	ctr	dif
6.9	6.4	0.5
7.6	6.7	0.9
7.3	5.4	1.9
7.6	8.2	-0.6
6.8	5.3	1.5
7.2	6.6	0.6
8.0	5.8	2.2
5.5	5.7	-0.2

For å illustrere, betrakter vi først våre målinger som parvise, dvs. observert på samme "objekt", f.eks. to matchede mødre. Differansen ($dif = trt - ctr$) noteres innen hvert par og ordnes i stigende rekkefølge ($rank$):

trt	ctr	dif	$rank$
7.6	8.2	-0.6	1
5.5	5.7	-0.2	2
6.9	6.4	0.5	3
7.2	6.6	0.6	4
7.6	6.7	0.9	5
6.8	5.3	1.5	6
7.3	5.4	1.9	7
8.0	5.8	2.2	8

Da datamaterialet er lite, kan det være fornuftig å *la være* å anta N-fordeling hvis vi ikke kan bekrefte dette på annen måte. Sentralgrenseteoremet holder ikke (se for eksempel flytskjema i Rosner s. 850-51). Vi kan imidlertid benytte såkalte *ikke-parametriske* metoder (egentlig *fordelings-uavhengige* metoder).

- I. **Sign test** (fortegnstest). Vi teller opp X , antall positive differanser, her: $X = 6$. Under H_0 om ingen endring vil X være binomisk fordelt med $n = 8$ og $p = 0,5$; dette gir $E(X) = np = n/2 = 4$.

$P(X \geq 6)$ finnes fra Rosner tabell 1 s. 818 ($n = 8$, $p = 0,5$) som $P(X = 6) + P(X = 7) + P(X = 8) = 0,11 + 0,03 + 0,003 \approx 0,14$. Tosidig verdi blir da $P \approx 0,28$.

En tosidig test vil forkaste H_0 om den observerte X er tilstrekkelig liten eller stor. Fra tabell 1 ser vi at verdiene 0 og 8 er de eneste som (tosidig) har sannsynlighet under 0,05 ($0,0039 \cdot 2 \approx 0,008$). For å forkaste H_0 på 0,05 nivå måtte vi altså enten observert $X = 0$ eller $X = 8$.

- II. **Wilcoxon's signed rank test.** Vi ordner nå observasjonene etter stigende *absoluttverdi* og tar både størrelse (rangorden, *rank*) og retning (fortegn) i betraktning. Bemerk beregning av gjennomsnittlig rangorden for sammenfallende verdier ("ties").

trt	ctr	dif	rank
5.5	5.7	-0.2	1
6.9	6.4	0.5	2
7.6	8.2	-0.6	3.5
7.2	6.6	0.6	3.5
7.6	6.7	0.9	5
6.8	5.3	1.5	6
7.3	5.4	1.9	7
8.0	5.8	2.2	8

Vi adderer nå rangordenen til de *positive* differansene ($2 + 3.5 + 5 + 6 + 7 + 8 = 31.5$). Under H_0 om ingen endring vil denne summen ha forventningsverdi $= n(n+1)/4 = 8*9/4 = 18$. Både lav og høy rangsum er usannsynlige under H_0 . Kritiske verdier finnes i Rosner tabell 11 s. 839. For en tosidig test med $n = 8$ og $\alpha = 0,05$ er disse verdiene (3, 33) *inklusive* (dette står i teksten s. 371). Verdien 31.5 er dermed ikke tilstrekkelig ekstrem og således "ikke signifikant". Vi må derfor konkludere med at behandlingen *ikke* hadde signifikant effekt på 0,05 nivå. [Eksakt $p = 0.063$].

La oss så anta at gruppene er uavhengige (ikke parvise), og ordner alle *originale* observasjoner i stigende rekkefølge, samtidig som vi holder rede på hvilken gruppe de tilhører:

wt	grp	rank
5.3	ctr	1
5.4	ctr	2
5.5	trt	3
5.7	ctr	4
5.8	ctr	5
6.4	ctr	6
6.6	ctr	7
6.7	ctr	8
6.8	trt	9
6.9	trt	10
7.2	trt	11
7.3	trt	12
7.6	trt	13.5
7.6	trt	13.5
8.0	trt	15
8.2	ctr	16

- III. **Wilcoxon's rank sum test** (ekvivalent: Mann-Whitney's U test). Avhengig av antallet i hver gruppe (det trenger ikke være likt!), vil rangsummen i hver av de to gruppene under H_0 om ingen forskjell bli "omtrent like stor" (det er opplagt hvis gruppene er like store). Vi adderer rangordenen i en av gruppene, f.eks. *trt* ($3 + 9 + 10 + 11 + 12 + 13.5 + 13.5 + 15 = 87$) og sammenlikner med kritiske verdier i Rosner tabell 12 s 840. For tosidig test med $\alpha = 0,05$ er disse verdiene (49, 87) *inklusive* (dette står i teksten s. 376). Vi kan altså forkaste H_0 på 0,05 nivå i dette tilfellet. [Eksakt $p = 0.047$]

Generelt vil tester som tar hensyn til både retning og størrelse (dvs. benytter mer informasjon) ha større styrke til å forkaste H_0 enn tester der bare retningen tas i betraktning. Testen i eks. I hadde mye mindre styrke ($p \approx 0,28$) enn de andre to. Vi fikk riktignok forskjellig konklusjon i eks. II og III pga. kravet til forkastning om (tosidig) $p < 0,05$; imidlertid ga begge likevel omtrent samme svar idet 0,063 ikke er vesentlig forskjellig fra 0,047.