Non-parametric methods

10.11.2009 Eirik Skogvoll

Rosner's chapter 9 about non-parametric methods is a bit difficult to follow, and the book almost exclusively employs normal approximation with a relatively large sample size. This may be a little confusing, because the methods are easiest to understand with small sample sizes.

Here is an excerpt from Rosner's table 8.18, p. 342, related to Problem 9.7, p. 380. The table reports the birth weight (`wt`, measured in pounds; 1 lb = 454 g) among children whose mothers have been taking some medication that possibly may delay premature delivery (`trt`); or 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

To illustrate, we first consider our observations as paired, i.e. on the same "object", e.g. two matched mothers. The difference (`dif = trt – ctr`) is noted within each pair, and ranked according to value.

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

As the sample size is small, it may be sensible not to assume a Normal distribution if this cannot be confirmed from other sources. The central limit theorem does not apply (see for example the flowchart in Rosner, pp. 850-51). We may however employ so called *non-parametric* methods (or really *distribution independent* methods).

I. **Sign test.** We count X, the number of positive differences, here: $X = 6$. Under H_0 of no change, X has a binomial distribution with $n = 8$ and $p = 0{,}5$; thus $E(X) = np = n/2 = 4$.

   $P(X \geq 6)$ is found in Rosner table 1 s. 818 (n = 8, p = 0,5) as $P(X = 6) + P(X = 7) + P(X = 8) = 0{,}11 + 0{,}03 + 0{,}003 \approx 0{,}14$. Two sided value is $P \approx 0{,}28$.

   A two tailed test will reject H_0 if observed X is sufficiently small or large. From table 1, we find that values 0 and 8 are the only values that (two tailed) receives a probability of less than 0,05 ($0{,}0039 * 2 \approx 0.008$). In order to reject H_0 at a 0,05 level we would have to observe either $X = 0$ or $X = 8$.

II. **Wilcoxon's signed rank test.** We now order our observations according to increasing *absolute value* and consider both rank and direction (sign). Note how an average rank is calculated in case of "ties" i.e. identical observations.

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

We now add the ranks of the *positive* differences (2 + 3,5 + 5 + 6 + 7 + 8 = 31,5). Under H_0 of no change, this sum has its expected value = n(n + 1)/4 = 8*9/4 = 18. Both small and large rank sums are unlikely under H_0. Critical values are found in Rosner table 11 p. 839. A two tailed test with n= 8 og $\alpha = 0,05$ has its critical values at (3, 33) *inclusive* (this is told in the text on p. 371). The observed value of 31.5 is therefore not sufficiently extreme and thus "not significant". We conclude that treatment was not effective at the 0,05 level. [Exact p = 0.063].

Let us now assume that the groups are independent (i.e. not date from a paired design). We rank all *original* observations according to magnitude, while keeping track of the group membership:

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** (equivalent: Mann-Whitney's U test). Depending on the number in each group (it doesn't have to be the same!), the rank sums in the two groups ought to be approximately equal if H_0 of no difference holds (this is obvious if group sizes are the same). We now add the ranks from one group, e.g. `trt` (3 + 9 + 10 + 11 + 12 + 13.5 + 13.5 + 15 = 87) and compare with critical values in Rosner table 12, p 840. For a two tailed test with $\alpha = 0,05$ the critical values are (49, 87) *inclusive* (this is told in the text on p. 376). We may thus reject H_0 at the 0,05 level [Exakt p = 0.047].

In general, tests that consider both direction and magnitude (i.e. utilizes more information) will have greater power to reject H_0 than tests that only consider direction. The test in example I had far less power (p-value $\approx$ 0,28) than the other two. We got different conclusions in Ex. II and III due to the requirement for rejection that two tailed p < 0,05. Still they both yielded approximately the same answer, as p = 0,063 is not substantially different from p= 0,047.