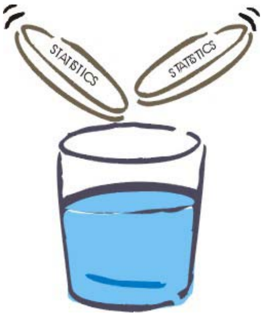
Statistics in Divided Doses



August 2004 No 7

Analysing data by non-parametric methods

Contents

- Non-parametric methods of analysis
- Wilcoxon signed rank sum test for analysing paired data
- Mann-Whitney test for analysing unpaired data

Non-parametric methods of analysis

What is meant by 'non-parametric'?

In earlier issues of *Statistics in Divided Doses* (*SiDD*) we used both the Normal and *t* distributions in hypothesis tests. In order for our testing to be valid, we assumed that our data (*protein X* levels) were from a population with an approximately Normal distribution. In real life, there will be occasions when we cannot make this assumption and we then have to use different types of hypothesis tests, which are variously known as distribution-free, rank, or non-parametric. None of these terms is entirely satisfactory but non-parametric is the one most commonly used. Such tests make (with some reservations) no assumptions concerning the distribution of test data and depend on the principle of data ranking.

What data are best suited to non-parametric analysis?

Skewed data (see *SiDD 2*) or data based on scores (eg symptom scores) are particularly suitable for analysis by non-parametric methods.

Could we use a non-parametric method to analyse data from a Normal distribution?

Yes, and it could be argued that it would be easier to use non-parametric tests routinely, especially with small samples where distributional characteristics cannot be checked. However, use of non-parametric tests involves some loss of statistical power compared to parametric methods, if the data are Normally distributed.

In practice, we usually choose a parametric method unless there is a clear indication that the underlying assumptions for Normality are not met.

What is meant by a rank and how is this calculated?

There are many examples of use of ranks in everyday life. Performance in a marathon is, for example, usually described in terms of place and timing; the competitor who gains first place (rank) is the one who completes the course in the shortest time. This simple system is identical with that used in statistical ranking i.e. ranking from the lowest value upwards.

The upper row of table 1 contains a selection of numbers; in the lower row the numbers have been ranked.

Table 1 - Numbers and ranks

Number	5	11	27	4	11	50
Rank	2	3.5	5	1	3.5	6

How has this been done and why are there decimal ranks?

If we rearrange the above numbers in ascending order, the process becomes more obvious (see table 2).

Table 2 - Numbers ranked in ascending order

Number	4	5	11	11	27	50
Rank	1	2	3.5	3.5	5	6

As the number 11 occurs twice, we say that it occupies a **tied rank**. In order to rank such identical numbers, we use a simple trick. Firstly, we work out the ranks the two numbers 11 would occupy if they were not the same. Clearly, these would be ranks 3 and 4. If we add these ranks together (7) and divide this by the number of ranks occupied by the identical numbers (2), we get 3.5. We therefore ascribe the rank of 3.5 to the two 11's. Remember that in doing this we have used up ranks 3 and 4 – the next rank that follows will then be 5.

Are there different types of non-parametric tests?

Yes. The two that we are looking at in this issue of *SiDD* are the **Wilcoxon signed rank sum test**, which is the non-parametric equivalent to the paired *t* test, and the **Mann-Whitney test,** the non-parametric equivalent of the unpaired *t* test.

The Wilcoxon signed rank sum test

The Wilcoxon signed rank sum test is a non-parametric method used to compare paired data or to compare values within a sample with a known reference value.

Remind me what is meant by paired data.

These are data that are derived from the same group of individuals. Measurement of clinical response in a sample of patients, before and after exposure to a trial drug, is an example of paired data.

NHS

Produced for healthcare professionals by North West Medicines Information Service,
The Pharmacy Practice Unit, 70 Pembroke Place, Liverpool, L69 3GF.
Editor: Frank Leach. Telephone: 0151 794 8117. E-mail: druginfo@liv.ac.uk

UKMi
UK Medicines Information

How is the Wilcoxon signed rank sum test applied?

The name gives us a useful clue. For a paired analysis, we first calculate the difference between each set of observations (e.g. before and after exposure to a test drug). We rank the differences in order of magnitude (ignoring the positive and negative signs), and then attribute a positive or negative sign to these ranks, depending on the direction of the difference between the 'before and after' values. The next step is to calculate the sum of all the positive and negative ranks. Referring to a table of the Wilcoxon signed rank sum test, we can then determine the probability level corresponding to our data. Confused? It's much easier to follow with a worked example!

An example of use of the Wilcoxon signed rank sum test.

Say we are interested in comparing two different methods of teaching statistics in the same group of 10 students. Method I uses conventional techniques while Method II employs games and puzzles. The statistical skills of the students are assessed after each course and the test marks compared (see table 3).

Table 3 – Statistics test marks (%) following two different teaching methods

No	Method I	Method II	Difference	Signed Rank
1	56	62	+ 6	+ 7
2	62	59	– 3	– 3.5
3	58	61	+ 3	+ 3.5
4	48	39	– 9	– 10
5	62	70	+ 8	+ 9
6	71	70	– 1	– 1
7	49	45	– 4	– 5
8	47	42	– 5	– 6
9	47	49	+ 2	+ 2
10	51	58	+ 7	+ 8

Sum of negative ranks = – 25.5
Sum of positive ranks = +29.5

The first three columns are (hopefully) self-explanatory. In column four we list the difference in exam results following the two teaching methods. In column five, we rank these differences and, after ranking, add a sign corresponding to the direction of the difference. For example, student 4 received 9 marks less after Method II than after Method 1. We record the difference as –9. Ignoring its minus sign, this is the largest difference among all the students and gains a rank of 10. Once ranked, we can now apply the minus sign, so that it joins the list of negative ranks.

What initial impressions do we gain from table 3?

The individual marks are reasonably close as are the differences between the results following each of the teaching methods. We suspect that Method II has not had much impact, but to see if these differences are statistically significant, we apply the Wilcoxon signed rank sum test.

What is the next step?

We calculate the sums of the positive and negative ranks separately. From table 3:

sum of negative ranks = –25.5
sum of positive ranks = +29.5

Can we check that our ranking is correct?

Yes, by means of the simple formula;
sum of ranks = n (n+1)/ 2 where n = number of ranks.
Substituting, we get 10 (10+1)/ 2 = 55 which is, ignoring signs, the sum of our ranks i.e. 25.5 + 29.5.

How do we use the Wilcoxon signed rank sum test table?

The section for an *n* value of 10 from the Wilcoxon signed rank sum test table is shown in table 4**.**

Table 4 – Section of Wilcoxon signed rank sum test table for n=10

n	Two-tailed probability (P)				
	0.2	0.1	0.05	0.02	0.01
10	14-41	10-45	8-47	5-50	3-52

You will see that as the P value decreases, the range of numbers in each column increases and that the numbers in each column all add up to 55 (since the sum of ranks when *n* =10 will always be 55).

To use the table, we look along the columns from the left to find the last column that does not contain the sum of our positive (or negative) ranks i.e. 29.5 or 25.5. Our values are contained in all the columns and we conclude that P>0.2. This clearly does not allow us to reject the null hypothesis. However, if our rank values had been 7 and 48 (positive or negative) the last column not including these would have been that for P=0.05. We could then conclude that the P value lay between 0.05 and 0.02 and that our results were statistically significant.

What is the overall conclusion from our example?

We have no evidence to suggest that the use of games and puzzles improves the learning of statistics.

Would the P values be the same if the rank sum signs were reversed?

Yes. If we had obtained rank sums of –29.5 and +25.5, the P value would be the same.

How can we use the Wilcoxon signed rank sum test to make a comparison with a standard value?

We might be interested to see, using a standard test, how the mathematical ability of our students compares with a fictitious average score of 35 points. Each student is tested and the difference between the student's score and the defined average is calculated. The set of differences is then analysed exactly as we did with our paired data. The results are set out in table 5.

Table 5 – Difference between students' scores and a standard average score of 35 points

Student number	Score in points	Difference from average (35)	Rank	Signed rank
1	20	– 15	4	– 4
2	22	– 13	2	– 2
3	15	– 20	7	– 7
4	19	– 16	5	– 5
5	25	– 10	1	– 1
6	21	– 14	3	– 3
7	7	– 28	10	– 10
8	8	– 27	9	– 9
9	18	– 17	6	– 6
10	11	– 24	8	– 8
Sum of negative ranks				**– 55**
Sum of positive ranks				**0**

The information contained in this bulletin is the best available from the resources at our disposal at the time of publication. Any opinions reflected are the author's own and may not necessarily reflect those of the Trust

Every student scored below the defined average; it looks like our students are having problems learning statistics. Putting this initial impression to the test, we look again at table 4, find that none of the columns contains our rank sum of 55, and conclude that P<0.01. We therefore reject the null hypothesis, conclude that our results are statistically significant and that our initial impression about our students' statistical abilities is probably justified.

Does the Wilcoxon method make any assumptions about the data?

The answer ought to be no because we have already said that non-parametric tests make no distributional assumptions. However, in practice, the data need to be approximately symmetric, although this is not an important restriction for a single sample test. In both our examples, the data were approximately symmetric without any extreme outliers.

What if the differences between paired data or a test and the standard reference value are zero?

These are ignored and the *n* value used in the Wilcoxon signed rank sum test table is decreased by the number of zero differences. For example, if one of the 10 students above had a test mark of 35, we would use an *n* value of 9 when referring to the test table.

The Mann Whitney test

The Mann-Whitney test is used for unpaired non-parametric data. These are data that are independent of each other e.g. a single set of results obtained from two distinct groups of patients.

How is the Mann Whitney test applied?

Using our earlier example, imagine that we have two groups of students following either a Method I or Method II statistics course. On this occasion, we compare the methods by means of one examination, the results of which are set out in table 6.

Table 6 – Statistics test marks (%) obtained by 20 students following different methods of teaching

Method I (n= 9)		Method II (n=11)	
Marks	**Rank**	**Marks**	**Rank**
		92	20
		91	19
		89	18
		87	17
82	16		
81	15		
		79	14
		76	13
74	12		
70	11		
		68	10
		67	9
65	8		
61	7		
		59	6
		57	5
40	4		
38	3		
30	2		
		20	1
Rank sum	**78**		**132**

In the table, we have ranked the results as though they were from a single sample. Method I (n=9) gives a total rank sum of 78 and Method II (n=11) a rank sum of 132. It is the smaller of the two totals that is used in the Mann Whitney test.

Can we check that our rank calculations are correct?

Yes, as before the sum of ranks = n (n+1)/2
i.e. (20 x 21)/2 = 210
Since 132 + 78 = 210, our calculations are correct.

What initial impressions do we gain from table 6?

The top four marks have gone to students following Method II. Thereafter, the distribution of marks is fairly even but some very poor marks have resulted following Method I.

How do we test the null hypothesis using the Mann Whitney test?

We use the Mann Whitney test table, the relevant section of which is reproduced below.

Table 7 – Section of Mann-Whitney test table

n_1	n_2	Two-tailed probability (P)				
		0.1	0.05	0.02	0.01	0.001
9	11	72 -117	68-121	63 -126	61 -128	53 -136

In our example, there were 9 students using Method I and their rank sum was 78. Using this number as our test statistic since it is the smallest of the two rank totals, we look along the columns to find the last column that does not include 78; we find that there is no column which answers this requirement. We therefore conclude that P>0.1 and that we cannot reject the null hypothesis.

What conclusion can we draw from our results?

We have no evidence that our two teaching methods differ in efficacy.

Can we use the Mann-Whitney test if there are tied ranks?

The test described is based on the assumption that there are no tied ranks. If there are many identical values we need to apply complicated corrections, which can be disregarded at this stage.

How can we describe the differences between the individual groups in our examples?

We can do this by estimating the respective median values. In table 6, for example, the median marks obtained using Method 1 and Method II are 65 and 76, respectively.

The information contained in this bulletin is the best available from the resources at our disposal at the time of publication. Any opinions reflected are the author's own and may not necessarily reflect those of the Trust