

A SAS/IML algorithm for exact nonparametric paired tests

Ein SAS/IML-Algorithmus für exakte nichtparametrische Tests für gepaarte Beobachtungen

Abstract

A common problem in practice is the comparison of two dependent samples. One possibility to evaluate such data is to compute the difference for each pair and apply a one-sample test. In this paper we discuss three nonparametric tests for the comparison of paired samples (i.e. one-sample tests). We present a macro written in SAS/IML to perform these tests as exact permutation tests. The macro is based on a shift-algorithm presented by Munzel & Brunner (2002) [1].

Zusammenfassung

Bei der Auswertung von Versuchen sind häufig zwei abhängige Stichproben zu vergleichen. Eine Möglichkeit der Auswertung besteht darin, für jedes Paar die Differenz zu berechnen und anschließend einen Einstichproben-Test auf diese Differenzen anzuwenden. In diesem Artikel werden drei verschiedene nichtparametrische Tests für den Vergleich gepaarter Daten (d.h. Einstichproben-Tests) diskutiert. Ein in SAS/IML geschriebenes Macro wird präsentiert, das diese Tests als exakte Permutationstests berechnet. Das Macro basiert auf einem von Munzel & Brunner (2002) [1] vorgestelltem Shift-Algorithmus.

Introduction

A common problem in practice is the comparison of two dependent samples. Examples are before/after comparisons, samples of the same subjects or samples of matched pairs of related subjects.

Here, we consider an example presented by Buck [2]. A study was performed to investigate an antibiotic. Its efficacy was measured by means of the number of leucocytes in the urine. The study population consisted of 10 subjects. Each subject was treated with a daily dose of 4 g. For each subject the number of leucocytes was determined before and 4 weeks after the treatment. The observed values for the 10 pairs are listed in Table 1.

The distribution of the number of leucocytes is unknown. Therefore parametric methods to determine the efficacy of the antibiotic, e.g. a one-sample t-test using the differences (baseline – after treatment), might be not appropriate. Instead, the differences (Table 1) can be analysed using nonparametric tests such as the Wilcoxon signed rank test, the modification of Wilcoxon's test according to Pratt, or a test based on the original data. In this article, we shall discuss all three tests and present a macro written in SAS/IML to perform these tests as exact permutation tests.

Ann-Kristin Leuchs¹
Markus Neuhäuser¹

¹ Department of Mathematics and Technique,
RheinAhrCampus, Koblenz
University of Applied
Sciences, Remagen,
Germany

Nonparametric tests for paired data

In the following we denote the two paired samples with $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$, thus the sample size is n . Moreover, let $d_i = x_i - y_i$ be the difference for the i -th pair (x_i, y_i) , $i = 1, \dots, n$.

One way to handle such paired data is to compute the difference of the two values for each pair. Then a test statistic based on these differences can be calculated. Due to the differences the two-sample problem is reduced to a one-sample problem.

For nonparametric tests there are often two alternative ways to carry out the test: one is to use the asymptotic distribution of the test statistic, the other is to perform an exact test, i.e. to use the exact null distribution of the statistic. In this paper we will focus primarily on the exact tests. To perform an exact nonparametric test, one needs to determine the exact null distribution (i.e. permutation distribution) of the test statistic.

To determine this distribution one needs to consider all possible permutations of the observed data. Since we consider paired data permutations are only possible within the pairs. Therefore the only way to permute is to swap the two values of a pair, i.e. to change the sign of the difference. Consequently, there are two possible permutations for one pair and 2^n possible permutations for n pairs. Given all these 2^n permutations the exact null

Table 1: Example: number of leucocytes/h

	Number of leucocytes/h (in millions)									
Leucocytes (baseline)	1.1	3.4	2.5	4.5	5.1	7.3	4.9	3.3	2.2	6.0
Leucocytes (after treatment)	0.3	0.4	0.2	0.2	0.3	2.8	4.9	0.5	4.2	6.0
Difference (baseline – after)	0.8	3.0	2.3	4.3	4.8	4.5	0	2.8	–2.0	0

distribution can be obtained by computing the test statistic for each permutation.

To be more precise, an exact nonparametric test for paired data consists of the following 4 steps irrespective of the used tests statistic [3].

1. The differences d_i have to be computed for the n pairs of data, and the test statistic has to be computed for the observed differences.
2. For the n pairs all 2^n possible assignments of plus and minus signs to the $|d_i|$'s have to be obtained.
3. The test statistic has to be computed for each of the 2^n possible assignments.
4. Then, the p-value can be computed as the proportion of assignments with a test statistic as or more supportive of the alternative than the observed value.

Below we will concentrate on three different tests: Wilcoxon signed rank test, Pratt's modification of Wilcoxon's signed rank test and a test based on original data. All three tests share the same null hypothesis, namely, the median θ of the differences is zero: $H_0: \theta = 0$. The alternative can be either one-sided ($H_1: \theta > 0$, respectively $H_1: \theta < 0$) or two-sided ($H_1: \theta \neq 0$). Moreover, we assume that the differences d_i are independent and symmetrically distributed. Note that the difference between two exchangeable random variables has a symmetric distribution [4].

Wilcoxon's signed rank test

At first assume that none of the differences is zero. The first step is to assign ranks to the absolute value of the differences $|d_i|$: the smallest $|d_i|$ gets the rank 1, the secondly smallest $|d_i|$ gets the rank 2 until the largest $|d_i|$ gets rank n . In the presence of ties (i.e. $|d_i| = |d_j|$ for some $i \neq j$) average ranks are assigned.

The statistic R^+ of the signed rank test is given by the sum of the ranks of the positive differences. One can compute the exact null distribution of this statistic using all 2^n permutations as mentioned above. The exact p-value either one-sided or two-sided can be obtained from this null distribution. When the asymptotic distribution of the test statistic is used, one needs the standardised statistic $(R^+ - E_0(R^+))/\text{Var}_0(R^+)$, i.e. the standardised rank sum of the positive differences, which is asymptotically standard normal. Under the null hypothesis the expected value of

the statistic is given by $E_0(R^+) = \frac{n(n+1)}{4}$. If there are no ties the variance under the null hypothesis is given by $\text{Var}_0(R^+) = \frac{n(n+1)(2n+1)}{24}$. In the presence of ties this variance

changes. A formula for the corrected variance is given by [5].

Up to now we assumed that none of the differences is zero. If there were differences equal to zero, Wilcoxon suggested discarding all zeros and applying the signed rank test to the reduced sample [6]. In most applications this suggestion is applied, as it is in our macro for the Wilcoxon signed rank test.

The Wilcoxon signed rank test can be performed using the SAS procedure UNIVARIATE. Even in SAS version 9.2 the test is, by default, carried out based on the asymptotic distribution if the remaining sample size is larger than 20. If the remaining sample size is ≤ 20 an exact test is performed.

Pratt's modification of Wilcoxon's signed rank test

In contrast to Wilcoxon's signed rank test where all zeros are ignored the modification according to Pratt [7] accounts for them. In the following n_0 denotes the number of zeros in the sample of differences. Pratt suggests assigning each zero the rank zero. All non-zero differences are ranked according to their absolute values with values from $n_0 + 1$ to n , i.e. the smallest non-zero difference $|d_i|$ gets rank $n_0 + 1$, etc., until the largest $|d_i|$ gets rank n . Analogous to Wilcoxon's signed rank test the statistic of Pratt's modification is given by the sum of the ranks for positive differences. One can compute the exact null distribution of this statistic using all 2^n permutations as mentioned above. The exact p-value either one-sided or two-sided can be obtained from this null distribution.

Note that due to the modification the expectation of the statistic under the null hypothesis changes:

$E_0(R^+) = \frac{n(n+1) - n_0(n_0+1)}{4}$. The variance is the same as for Wilcoxon's test. The standardized test statistic is still asymptotic standard normal distributed [8]. The test according to Pratt is not implemented in SAS.

Test based on original data

Wilcoxon's signed rank test as well as the modification according to Pratt is based on ranks. However, the null hypothesis that the median of the differences is zero can also be tested with a test based on the original data. Instead of assigning ranks and summing up the ranks of positive differences, for this test the statistic is computed by summing up the positive differences. Thus, this test uses the information about the sign of the difference and

also the magnitude of the difference. Analogous to Wilcoxon's signed rank test and Pratt's modification one can compute the exact null distribution of this statistic using all 2^n permutations as mentioned above. The exact p-value either one-sided or two-sided can be read from this null distribution. Note that several alternative forms of the statistic are also possible [3].

This permutation test is implemented in StatXact (Cytel Software Corporation, Cambridge, Mass.), but is not available within a SAS procedure. Note that StatXact also offers the Wilcoxon signed rank test.

Comparison and discussion of the tests

So far we only considered the null hypothesis $H_0: \theta = 0$. Note that all tests are likewise applicable if the null hypothesis reads $H_0: \theta = \theta_0$. In this case one only needs to subtract θ_0 from the differences and then apply one of the presented tests.

As we already mentioned Wilcoxon's signed rank tests drops zeros. However, ignoring zeros may lead to contradictory tests results [7]. Namely, one sample is not significantly positive while a more negative sample (all observations where reduced equally in magnitude) can be significantly positive. For illustration the following sample is considered

0, 2, 3, 4, 6, 7, 8, 9, 11, 14, 15, 17, -18,

as presented by Pratt [7]. The null hypothesis to be tested is $H_0: \theta > 0$. To apply Wilcoxon's signed rank test the zero is discarded and then ranks are assigned. This leads to the signed ranks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, -12. The observed sum of positive ranks is 66, the p-value of Wilcoxon's signed rank test is $60/2^{12} = 0.0171$.

Now consider the decreased sample -0.5, 1.5, 2.5, 3.5, 5.5, 6.5, 7.5, 8.5, 10.5, 13.5, 14.5, 16.5, -17.5 with signed ranks -1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, -13. For this sample the sum of positive ranks is 77 and the p-value is $109/2^{13} = 0.0133$. Thus the first sample has the larger p-value, indicating that it is less significantly positive. In this example we decreased the sample by an amount of 0.5. One gets the same result for any amount less than 1.

Note that such a contradictory result is possible in the other direction, too. There are examples where one sample is significantly positive while an increased sample is not significantly positive. For the modification according to Pratt which accounts for zeros this problem does not occur. In the example mentioned above the p-value of Pratt's test is 0.0120 for the first sample. Obviously, the p-value of the decreased sample (0.0133) does not change. For more details it is referred to Pratt [7]. All p-values in this paragraph were calculated with the new SAS-Macro presented in this paper.

To compute the exact null distributions of the three statistics mentioned above we used a shift-algorithm presented by Munzel & Brunner [1] (originally developed by Streitberg & Röhmle [9]). The main part of their paper

was the presentation of a new nonparametric test for paired ordered categorical data. For such data none of the three tests (Wilcoxon's signed rank test, the modification according to Pratt and the test based on original data) should be used since for them differences need to be computed. For more details it is referred to Munzel & Brunner [1].

For applications the question arises when to use which test. For ordered categorical data the test presented by Munzel & Brunner [1] can be recommended since the other three tests discussed here cannot be applied. However, the sign test [3] is an alternative which can be applied for ordered categorical data.

For the other tests the null hypothesis states that the population median of the difference is zero. For this test problem ignoring zero differences is not appropriate [10]. Therefore, we suggest Pratt's modification over Wilcoxon's signed rank test whenever zero differences are possible. A different question is whether a rank test such as Pratt's modification or a test based on original data should be applied. The latter may be preferable when the underlying data are approximately normal. However, for non-normal distributions rank tests are relatively powerful. Therefore, rank permutation tests are still useful although more complicated permutation tests can be carried out with modern PCs [11].

Example

Consider again the example presented above in Table 1. To evaluate the data using Wilcoxon's signed rank test all zeros are dropped and then ranks need to be assigned. Therefore the absolute values of the differences (last row of Table 1) need to be ordered and the smallest value gets rank 1, the secondly smallest value gets rank 2, etc. The ranks including the signs are displayed in Table 2. The tests statistic of Wilcoxon's signed rank test is the sum of the positive ranks which is 34.

To compute the p-values of the exact tests all $2^8 = 256$ possible permutations need to be considered and for each permutation the statistic needs to be computed. We obtain the p-value for the one-sided alternative "median greater than zero" as the probability of those statistics which are at least 34 (observed statistic). The observed permutation contains only one negative value which is the secondly smallest rank. Consequently, there are overall three permutations with a statistic of at least 34, namely, the observed permutation and the two permutations with only positive ranks except the smallest one which is either positive or negative. Therefore the one-sided p-value is given by $3/256 = 0.0117$.

When considering the two-sided alternative, permutations with a statistic of at most 2 are as supportive of the alternative as permutations with a statistic of at least 34, because 2 and 34 have the same distance to 18 which is the expected value of the rank sum under the null hypothesis. Multiplying the permutations with a statistic of at least 34 with -1 yield the permutations with a statistic

Table 2: Example: ranks for Wilcoxon's signed rank test and the modification according to Pratt

Difference (baseline – after)	Number of leucocytes/h									
	0.8	3.0	2.3	4.3	4.8	4.5	0	2.8	–2.0	0
	Ranks									
Wilcoxon's signed rank test	1	5	3	6	8	7	/	4	–2	/
Modification according to Pratt	3	7	5	8	10	9	0	6	–4	0

of at most 2. Therefore the two-sided p-value is given by $6/256 = 0.0234$.

Wilcoxon's signed rank test simply ignores the zeros. In contrast Pratt's modification accounts for them. All zeros are assigned the rank zero and for all other differences ranks are assigned as with Wilcoxon except for the fact, that the smallest value of the non-zero differences gets the rank $n_0 + 1$, the next one $n_0 + 2$, etc. where n_0 is the number of zeros. The ranks according to Pratt's modification are displayed in Table 2. As for the signed rank test the statistic for Pratt is the sum of the positive ranks which is 48 in this example. The one-sided p-value (alternative: "median greater than zero") is given by the probability of the permutations with a statistic of at least 48: $3/256 = 0.0117$.

The expected value of the test statistic is 26. Therefore the two-sided p-value is $6/256 = 0.0234$ (permutations with statistics that are at least 48 and at most 4). Note that for Pratt's modification the number of permutations is 2^8 as well, because one cannot assign a sign to the zeros.

For the test based on original data the observed differences were used (no ranks need to be assigned). The tests statistic which is the sum of the positive differences is 22.5. The permutation distribution leads the one-sided p-value $3/256 = 0.0117$ and the two-sided p-value $6/256 = 0.0234$. Note that, under the null hypothesis, the expected value of the statistic is 12.25.

This example yields the same p-values for all three tests. Note that there are examples where the p-values differ.

The SAS-Macro for exact paired tests

We will now present a macro written in SAS/IML using a shift-algorithm [1] to perform the three presented tests. Note that the test introduced by Munzel & Brunner [1] can also be carried out with the algorithm presented in [1]. Besides the computation of ranks one important part of the program is the generation of the permutation null distribution. For all discussed tests the procedure to compute this distribution is identical. An observed data vector d is given which contains the signed ranks of the differences (for Wilcoxon's signed rank test and the modification according to Pratt) or the original observed differences (for the test based on original data). The statistic is the sum of the positive values in this vector irrespective of the test and can be written as $s' d$ where s' is the transpose of the vector $s \in \{0,1\}^n$.

To get the permutation null distribution the probabilities $g(t)/2^n$ for all possible values t of the statistic need to be determined, where $g(t)$ is the number of permutations which leads to a specific statistic t (i.e. $g(t)$ is the number of elements in $\{s' d = t \mid s \in \{0,1\}^n\}$).

Shift-Algorithm

The shift-algorithm presented by Munzel & Brunner [1] is an easy way to compute these probabilities or rather the numbers $g(t)$, but only for a vector d of integer values. Since in our case the vector d can contain decimal numbers (e.g. average ranks), we multiply the data vector d with 10^k for a sufficiently high value k , then we apply the shift-algorithm and finally divide by 10^k .

In the following the shift-algorithm is explained by means of the vector $d = (d_1, d_2, d_3) = (2, -4, 4)$ (i.e. $n = 3$). First Munzel & Brunner divide the numbers in d by their largest common factor so that the computational capacity is reduced. This leads to $d_{\text{mod}} = (1, -2, 2)$. Table 3 illustrates the shift-algorithm for the modified vector d_{mod} . The first column contains all integer values between the lowest possible value of the statistic t which is 0 and the biggest one which is $1 + 2 + 2 = 5$.

Note, that only one permutation, namely "all elements in d are negative", yields a statistic $t = 0$. Therefore the algorithm starts with the vector $(1, 0, 0, 0, 0)$ which finally will contain the numbers $g(t)$: i.e. at the beginning only the permutation "all elements are 0" is considered, which leads to the statistic $t = 0$.

Step 1

This starting vector is now shifted by the absolute value of $d_{\text{mod},1} = 1$. The result $(0, 1, 0, 0, 0)$ (column 3 of Table 3) corresponds to the situation that $d_{\text{mod},1}$ is positive and all other observations are negative, i.e. $t = 1$. The outcome of the first step (column 4) which is the sum of the starting vector and the shifted vector contains the numbers $g(t)$, if only the permutations where $d_{\text{mod},2}$ and $d_{\text{mod},3}$ are negative are considered, i.e. there is one possibility to get the statistic $t = 0$, if $d_{\text{mod},1}$ is negative and $t = 1$ if $d_{\text{mod},1}$ is positive.

Step 2

The result of step 1 is now shifted by the absolute value $|d_{\text{mod},2}| = 2$. This shifted vector (column 5) corresponds to the case that $d_{\text{mod},1}$ has an arbitrary sign, $d_{\text{mod},2}$ is positive

Table 3: Shift-algorithm for (1, -2, 2)

Possible values t	Start	Shift by $ d_{\text{mod},1} = 1$	Result step 1	Shift by $ d_{\text{mod},2} = 2$	Result step 2	Shift by $ d_{\text{mod},3} = 2$	Result step 3 ($= g(t)$)	Prob.
0	1	0	1	0	1	0	1	0.125
1	0	1	1	0	1	0	1	0.125
2	0	0	0	1	1	1	2	0.250
3	0	0	0	1	1	1	2	0.250
4	0	0	0	0	0	1	1	0.125
5	0	0	0	0	0	1	1	0.125

and $d_{\text{mod},3}$ is negative. Therefore the sum of column 4 and 5 (column 6) contains the numbers $g(t)$, if only the permutations, where $d_{\text{mod},3}$ is negative, are considered.

Step 3

Finally the resulting vector of step 2 is shifted by $|d_{\text{mod},3}| = 2$ (column 7). The shifted vector contains the numbers $g(t)$, if only the permutations, where $|d_{\text{mod},3}|$ is positive, are considered. If this vector is summed up with the result of step 2, it leads to the numbers $g(t)$ to be determined: i.e. (1,1,2,2,1,1) are the numbers of all $2^n = 8$ permutations which lead to the statistics (0,1,2,3,4,5). We get the corresponding probabilities by dividing the numbers $g(t)$ by 2^n (column 9).

Table 3 shows the permutation-distribution of d_{mod} as a result of the shift-algorithm. The aim was to determine the permutation distribution of the original vector d . This can be easily done by multiplying the first column of Table 3 with the largest common factor of the values in d . The resulting distribution of d is displayed in Table 4.

Table 4: Permutation distribution for (2, -4, 4)

Statistic	Probability
0	0.125
2	0.125
4	0.250
6	0.250
8	0.125
10	0.125

The first parameter *data* specifies the dataset. The second parameter *label_diff* is the name of the variable inside the dataset *data* which should be analysed. The third parameter *test* specifies the test to be performed. You can choose 'signed' for Wilcoxon's signed rank test, 'pratt' for the modification according to Pratt and 'original' for the test based on original data.

The parameter *round* is an optional parameter which is 4 by default. This parameter specifies the number of decimal places to remain after rounding. Rounding is only performed if the test based on original data is computed. Note that in the long version of the macro (Attachment 2) you can download on the journal's homepage there is yet another parameter *alternative*. With this parameter the alternative is specified. Choosing *alternative* = 'two' leads to the output of a two-sided p-value. Choosing *alternative* = 'greater' or 'less' leads to the output of a one-sided p-value. Finally, choosing *alternative* = 'all' leads to the output of all three p-values.

Let us again consider the example of Buck [2]. To use the macro the 10 values of the difference are saved in a data set.

```
DATA example;
INPUT diff @@;
CARDS;
0.8 3 2.3 4.3 4.8 4.5 0 2.8 -2 0
;
RUN;
```

Then the macro is invoked.

```
%signedrank(example, diff, 'pratt');
```

The output is presented here for Pratt's test. The total sample size, the number of non-zeros, the observed value of the test statistic, and the p-value(s) are given (Table 5). As you can see, we get the same two-sided p-value (0.0234) as above.

Application of the macro

A short version of the macro which will be explained below is printed in the appendix (Attachment 1) and can be downloaded at the journal's homepage. In order to use the macro properly the following remarks on the parameters are necessary.

As one can see the MACRO is called *signedrank* and three parameters need to be specified

```
%MACRO signedrank(data, label_diff, test, round=4);
```

Table 5

test	n	n (nonzero)	statistic	p-value (less)	p-value (greater)	p-value (two-sided)
pratt	10	8	48	0.9921875	0.0117188	0.0234375

Notes

Acknowledgement

The authors gratefully acknowledge support of this work by the Ministry for Education, Science, Youth, and Culture of Rhineland-Palatinate for a Competence Center in Bio-mathematics.

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from

<http://www.egms.de/en/journals/mibe/2010-6/mibe000104.shtml>

1. MIBE-Macro-Short-Version.pdf (312 KB)
Appendix: Short version of the macro
2. MIBE-Macro-Long-Version_korr.pdf (287 KB)
Long version of the macro

References

1. Munzel U, Brunner E. An exact paired rank test. *Biom J.* 2002;44(5):584-93. DOI: 10.1002/1521-4036(200207)44:5<584::AID-BIMJ584>3.0.CO;2-9
2. Buck W. Der Vorzeichen-Rang-Test nach Pratt. *Meth Inf Med.* 1975;14(4):224-30.
3. Higgins JJ. An introduction to modern nonparametric statistics. Brooks/Cole: Pacific Grove; 2004.
4. Randles RH, Wolfe DA. Introduction to the theory of nonparametric statistics. Wiley: New York; 1979.
5. Hollander M, Wolfe DA. Nonparametric statistical methods. 2nd ed. Wiley: New York; 1999.
6. Wilcoxon F. Some rapid approximate statistical procedures. New York: American Cyanamid Co; 1949.
7. Pratt JW. Remarks on zeros and ties in the Wilcoxon signed rank procedures. *J Am Stat Assoc.* 1959;54:655-67.

8. Buck W. Signed-rank tests in the presence of ties (with extended tables). *Biom J.* 1979;21(6):501-26. DOI: 10.1002/bimj.4710210602
9. Streitberg B, Röhmel J. Exakte Verteilungen für Rang- und Randomisierungstests im allgemeinen c-Stichprobenfall. *EDV Biol Med.* 1987;18:12-19.
10. Larocque D, Randles RH. Confidence intervals for a discrete population median. *Am Stat.* 2008;62(1):32-9. DOI: 10.1198/000313008X269738
11. Neuhaus M. Efficiency comparisons of rank and permutation tests. *Stat Med.* 2005;24(11):1777-8. DOI: 10.1002/sim.1939

Erratum

A minus sign has been added in Attach. 2 Long version of the macro, page 2:

```
shift = j(adshort[k],1,0) // prob[1:lng+1
      -adshort[k]];
```

Corresponding author:

Ann-Kristin Leuchs
Department of Mathematics and Technique,
RheinAhrCampus, Koblenz University of Applied Sciences,
Südallee 2, 53424 Remagen, Germany
aleuchs@rheinahrcampus.de

Please cite as

Leuchs AK, Neuhaus M. A SAS/IML algorithm for exact nonparametric paired tests. *GMS Med Inform Biom Epidemiol.* 2010;6(1):Doc04. DOI: 10.3205/mibe000104, URN: urn:nbn:de:0183-mibe0001041

This article is freely available from

<http://www.egms.de/en/journals/mibe/2010-6/mibe000104.shtml>

Published: 2010-12-15

Published with erratum: 2012-04-05

Copyright

©2010 Leuchs et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share — to copy, distribute and transmit the work, provided the original author and source are credited.