

Exact Methods in the NPAR1WAY Procedure

Anupama Narayanan and Donna Watts, SAS Institute Inc., Cary, NC

Abstract

Exact nonparametric methods have an advantage over asymptotic methods since they remain valid for very small sample sizes, as well as for data that are sparse, skewed, or heavily tied. However, computing exact p -values by direct enumeration can be very time-consuming, and may be infeasible for many problems. Over the past few years, many new algorithms for exact computations have been published. Beginning with Release 6.11 of the SAS® System, exact p -values are available in the NPAR1WAY procedure. This procedure uses Mehta and Patel's network algorithm, substantially reducing the computational effort required.

The NPAR1WAY procedure computes exact p -values for the simple linear rank statistics based on Wilcoxon scores, median scores, Van der Waerden scores, and Savage scores. Exact p -values are computed for these statistics when the data are classified into two levels (two-sample tests), and when the data are classified into more than two levels (multi-sample tests). For the two-sample case, the simple linear rank statistics listed above correspond to the Wilcoxon-Mann-Whitney, Median, Van der Waerden, and Savage tests, respectively. For the multi-sample case, the NPAR1WAY procedure gives exact p -values for the Kruskal-Wallis, Brown-Mood, k -sample Van der Waerden, and k -sample Savage tests.

Introduction

Many clinical trials deal with the comparison of populations of subjects whose responses are largely categorical in nature. Historically, statistical inference for such studies involved large-sample approximations. However, sample sizes for these studies are not always large enough to justify the use of asymptotic methods. Also, the asymptotic methods may be inadequate when data are sparse, skewed, or heavily tied. Exact inference remains valid in such situations and, hence, is a preferred approach.

R. A. Fisher (1934) introduced conditional exact methods for small samples. This approach eliminates nuisance parameters by conditioning on functions of the observed values. Exact nonparametric methods are most commonly based on this approach.

Nonparametric methods are also known as *distribution-free methods* since they do not assume an underlying parametric distribution (such as normal) for the data. These methods are particularly useful when the underlying distribution is nonstandard or unknown, or when data consist of ranks (and not values) of random variables. The next section discusses some nonparametric tests offered in the

NPAR1WAY procedure. The sections that follow describe the new EXACT statement in the NPAR1WAY procedure and the exact p -value computations and illustrate their applications with examples derived from clinical trials.

The EXACT statement is also available in the NPAR1WAY procedure in Release 6.09E of the SAS System for mainframes and minicomputers.

Nonparametric Tests

This section briefly reviews test statistics for nonparametric tests based on ranks. For details on these and other tests, refer to Hollander and Wolfe (1973), Hajek (1969) and "Simple Linear Rank Statistics" in the SAS/STAT® *User's Guide* under the NPAR1WAY procedure.

Suppose the data consist of N observations that are classified into two levels: n_1 observations in the first population with distribution function F_1 and n_2 observations in the second population with distribution function F_2 .

The aim is to compare the two distributions for a shift in location. In other words, to test the hypothesis
 $H_0: F_1(w) = F_2(w)$ for all w vs
 $H_1: F_1(w - \theta) = F_2(w)$ for all w , or vs
 $H_2: F_1(w + \theta) = F_2(w)$ for all w , for an arbitrary shift parameter θ

Consider a random sample from each of the two populations: X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} . To perform these nonparametric tests, combine and arrange these N observations in an increasing order and assign rank R_j to the j th observation. If k observations are tied at a value, assign the mean of k consecutive ranks to each observation.

Assign scores w_j to the observations based on these ranks, and compute a statistic using these scores. In the case of two-sample tests, the statistic has a linear form and is computed as

$$S = \sum_{j=1}^N c_j w_j$$

where c_j is an indicator denoting the class (0 or 1) to which the j th observation belongs.

Wilcoxon-Mann-Whitney Scores

These scores are the ranks assigned to the observations using the method described above.

$$w_j = R_j$$

These scores are used to test for location shifts in logistic distributions.

Median Scores

These scores are 1 for points above the median, 0 otherwise.

$$w_j = \begin{cases} 1 & \text{if } R_j > (N + 1)/2 \\ 0 & \text{otherwise} \end{cases}$$

Median scores are used with double exponential distributions.

Van der Waerden Scores

These scores are approximations of the expected values of the order statistics for a normal distribution.

$$w_j = \Phi^{-1}(R_j/(N + 1))$$

where Φ is the distribution function for normal distribution. These scores are used with normal distributions.

Savage Scores

These scores are expected values of order statistics for the exponential distribution, with 1 subtracted to center the scores around 0.

$$w_j = \sum_{i=1}^{R_j} \left[\frac{1}{N - i + 1} \right] - 1$$

Savage scores are used to compare scale differences in exponential distributions or location shifts in extreme value distributions.

In the case of multi-sample tests, the statistic is of a different form (not linear). Suppose the data consist of N observations, and they are classified into $k > 2$ levels with n_i observations in level i . Let X_{ij} denote the j th observed value in the i th level.

To test the null hypothesis that the underlying distributions are the same for all levels against an alternative that not all distributions are the same, rank all the N observations in increasing order and assign ranks as described in the two-sample case. Let R_{ij} be the rank of X_{ij} .

Kruskal-Wallis Statistic

This statistic has the form

$$S = \left(\frac{12}{N(N + 1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N + 1)$$

where $R_i = \sum_{j=1}^{n_i} R_{ij}$ is the sum of ranks for the i th level.

Other multi-sample scores are similar to their two-sample equivalent scores. For details, refer to Hajek (1969).

Algorithms for Exact Nonparametric Tests

Several algorithms have been published for computation of exact p -values. Agresti (1992) reviews these algorithms in detail. Algorithms based on direct enumeration can be very time-consuming and are feasible only for small problems. An algorithm, known as the network algorithm, developed by Mehta and Patel (1983) provides a substantial advantage over direct enumeration. It requires significantly less time for computations than the direct enumeration methods, as

shown by Mehta and Patel (1983) and Mehta et al. (1984, 1991). The network algorithm has been applied to several problems, as discussed by Mehta and Patel (1983, 1986), Mehta et al. (1984, 1985, 1991), Hirji et al. (1987), and Agresti et al. (1990). The NPAR1WAY procedure uses the network algorithm as described in Mehta and Patel (1983), Mehta et al. (1984, 1991), and Agresti et al. (1990). It also uses bounds described in Valz and Thompson (1994).

The EXACT Statement

The NPAR1WAY procedure can be invoked using the following statements:

```
PROC NPAR1WAY <options>;
CLASS variable;
VAR variable;
EXACT <options>;
```

The PROC NPAR1WAY statement invokes the NPAR1WAY procedure and optionally identifies a data set or requests particular analyses. The keywords WILCOXON, MEDIAN, VW and SAVAGE can be used to compute linear rank statistics based on Wilcoxon scores, Median scores, Van der Waerden scores, and Savage scores. Asymptotic p -values are computed for the tests specified in the PROC NPAR1WAY statement. The CLASS statement specifies exactly one classification variable. This variable identifies groups in the data. The VAR statement can be used to denote the response variable.

The new EXACT statement specifies the statistics for which you want exact p -values. The keywords WILCOXON, MEDIAN, VW and SAVAGE specify the scores that are used to compute the linear rank statistic. Exact p -values are available for these statistics when data are classified into two levels (two-sample tests) or more (multi-sample tests). For two-sample tests these keywords correspond to Wilcoxon-Mann-Whitney, median, Van der Waerden and Savage scores respectively. For multi-sample tests, they correspond to Kruskal-Wallis, Brown-Mood, k -sample Van der Waerden and k -sample Savage scores respectively.

If no keywords are listed in the EXACT statement, then exact p -values are computed for all the linear rank statistics requested in the PROC NPAR1WAY statement.

Exact p -values

If exact analysis is requested, the NPAR1WAY procedure computes exact p -values only when two or more levels of the CLASS variable are present.

When there are two groups (the data are classified into two levels), the NPAR1WAY procedure gives exact one-sided and two-sided p -values for each statistic specified in the EXACT statement. The exact one-sided p -value is printed as the right-sided p -value when the statistic is greater than its mean. Otherwise, the left-sided p -value is printed. Mathematically, the right- and left-sided p -values can be expressed as

$$p_1 = P(\text{Rank Statistic} \geq S) \quad \text{if } S > \text{Mean}$$

$$p_1 = P(\text{Rank Statistic} \leq S) \quad \text{if } S \leq \text{Mean}$$

where S is the observed rank statistic, "Mean" is the expected value of the test statistic, and p_1 is the one-sided p -value.

The NPAR1WAY procedure computes the two-sided p -value as the sum of the one-sided p -value and the corresponding area in the opposite tail of the distribution of the statistic, equidistant from the mean. Mathematically, the two-sided p -value, p_2 can be expressed as

$$p_2 = P(|\text{Rank Statistic} - \text{Mean}| \geq |S - \text{Mean}|)$$

If there are more than two levels and the EXACT statement is specified, the NPAR1WAY procedure prints only the two-sided p -value. (A one-sided p -value does not exist, since the tests are inherently two-sided.)

Examples

Two-Sample Data: Wilcoxon Test

Researchers conducted an experiment to compare the effects of two stimulants. Thirteen randomly selected subjects received the first stimulant, and six randomly selected subjects received the second stimulant. The reaction times (in minutes) were measured while the subjects were under the influence of the stimulants.

The following SAS statements create the data set REACT, which contains the observed reaction times for each stimulant. The variable STIM represents Stimulant 1 or 2. The variable TIME represents the reaction times observed for subjects receiving one of the two stimulants.

```
data react;
  input stim time @@;
  cards;
  1 1.94 1 1.94 1 2.92 1 2.92
  1 2.92 1 2.92 1 3.27 1 3.27
  1 3.27 1 3.27 1 3.70 1 3.70
  1 3.74 2 3.27 2 3.27 2 3.27
  2 3.70 2 3.70 2 3.74
  ;
```

The NPAR1WAY procedure tests the null hypothesis of no difference between the effects of the two stimulants against the alternative that stimulant 1 has smaller reaction times than stimulant 2. The WILCOXON option specifies that Wilcoxon scores are to be used in the computations. The CLASS statement specifies that the two classes are determined by the variable STIM. The VAR statement identifies TIME as the response variable. The EXACT option is used to request exact p -values.

```
proc npar1way wilcoxon data=react;
  class stim;
  var time;
  exact;
run;
```

Output 1. Results of the Test on Stimulants

N P A R 1 W A Y P R O C E D U R E					
Wilcoxon Scores (Rank Sums) for Variable TIME Classified by Variable STIM					
STIM	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	13	110.500000	130.0	11.0047836	8.5000000
2	6	79.500000	60.0	11.0047836	13.2500000
Average Scores Were Used for Ties					
Wilcoxon 2-Sample Test			S = 79.5000		
Exact P-Values					
(One-sided) Prob >= S = 0.0527					
(Two-sided) Prob >= S - Mean = 0.1054					
Normal Approximation (with Continuity Correction of .5)					
Z = 1.72652 Prob > Z = 0.0843					
T-Test Approx. Significance = 0.1014					
Kruskal-Wallis Test (Chi-Square Approximation)					
CHISQ = 3.1398 DF = 1 Prob > CHISQ = 0.0764					

Output 1 displays the results of testing the null hypothesis. The one-sided exact p -value of 0.0527 is not significant at the $\alpha = 0.05$ level. However, the normal approximation yields a one-sided p -value of 0.0422 (this is half of the p -value shown in Output 1 since the normal distribution is symmetric). This p -value is significant at the $\alpha = 0.05$ level. Thus, the exact p -value may result in nonrejection of the null hypothesis at the 0.05 level of significance, while the normal approximation method may not. Such discrepancies are likely to occur when the samples are small or data are sparse.

Multi-Sample Data: Savage Test

A researcher conducting a laboratory experiment randomly assigned fifteen mice to receive one of three drugs. The survival time (in days) was then recorded.

The following SAS statements create the data set MICE, which contains the observed survival times for all the mice. The variable TRT denotes the treatment and the variable DAYS denotes the number of days a mouse survived after receiving the treatment.

```
data mice;
  input trt $ days @@;
  cards;
  1 1 1 1 3 1 3 1 4
  2 3 2 4 2 4 2 4 2 15
  3 4 3 4 3 10 3 10 3 26
  ;
```

The NPAR1WAY procedure is used to test the null hypothesis that there is no difference in the survival times among the three drugs. The SAVAGE option specifies that the Savage scores are to be used. The variable TRT is the CLASS variable, and the VAR statement specifies that the variable DAYS is the response variable.

```
proc npar1way savage data=mice;
  class trt;
  var days;
  exact;
run;
```

Output 2. Results of Savage Test on Multi-Sample Data

```
The SAS System 1
NPAR1WAY PROCEDURE
Savage Scores (Exponential) for Variable DAYS
Classified by Variable TRT
```

TRT	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	5	-3.36798017	0.0	1.63455487	-.673596034
2	5	0.09561827	0.0	1.63455487	0.019123654
3	5	3.27236190	0.0	1.63455487	0.654472379

Average Scores Were Used for Ties

Savage 1-Way Analysis S = 5.5047

Exact P-Value Prob >= S = 0.0445
 Chi-Square Approximation DF = 2 Prob > S = 0.0638

The results of the Savage test are shown in Output 2. The exact p -value is 0.0445, which is significant at the $\alpha = 0.05$ level. However, the p -value based on the chi-square approximation is 0.0638, which results in nonrejection of the null hypothesis at $\alpha = 0.05$ level.

Two-Sample Data: Wilcoxon Test

Fifty-nine female patients with rheumatoid arthritis who participated in a clinical trial were assigned to two groups, active and placebo. The response status (excellent, good, moderate, fair, poor) of each patient was recorded.

The following SAS statements create the data set ARTH, which contains the observed status values for all the patients. The variable TRT denotes the treatment received and the variable RESP denotes the response status of a patient.

```
data one;
  input trt $ resp freq @@;
  cards;
Active 1 5 Active 2 11 Active 3 5
Active 4 1 Active 5 5 Placebo 1 2
Placebo 2 4 Placebo 3 7 Placebo 4 7
Placebo 5 12
;
data arth;
  set one;
  do i=1 to freq;
    output arth;
  end;
run;
```

The NPAR1WAY procedure is used to test the null hypothesis that there is no difference in patient response status. The WILCOXON option specifies that Wilcoxon scores are to be used in the computations. The variable TRT is the CLASS variable, and the VAR statement specifies that the variable RESP is the response variable.

```
proc npar1way wilcoxon data=arth;
  class trt;
  var resp;
  exact;
run;
```

Output 3. Results of Wilcoxon Test on Two-Sample Data

```
The SAS System 1
NPAR1WAY PROCEDURE
Wilcoxon Scores (Rank Sums) for Variable RESP
Classified by Variable TRT
```

TRT	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Active	27	621.0	810.0	63.9727441	23.0000000
Placebo	32	1149.0	960.0	63.9727441	35.9062500

Average Scores Were Used for Ties

Wilcoxon 2-Sample Test S = 621.000

Exact P-Values
 (One-sided) Prob <= S = 0.0014
 (Two-sided) Prob >= |S - Mean| = 0.0028

Normal Approximation (with Continuity Correction of .5)
 Z = -2.94657 Prob > |Z| = 0.0032

T-Test Approx. Significance = 0.0046

Kruskal-Wallis Test (Chi-Square Approximation)
 CHISQ = 8.7284 DF = 1 Prob > CHISQ = 0.0031

The results of testing the null hypothesis are shown in Output 3. The p -values obtained from the exact test and normal approximation are 0.0028 and 0.0032, respectively. Both p -values are significant at the $\alpha = 0.01$ level. In this case, the asymptotic p -value (based on the normal approximation) is more conservative than the exact p -value.

Computational Resources

Although the computational algorithm is fast, the computational time for exact p -values can still be prohibitive, depending on the number of groups, the number of distinct response variables, the total sample size, and the speed and memory available on your computer.

If you do not wish to wait for the exact computations to complete, you can terminate exact computations and exit to the NPAR1WAY procedure at any time by pressing the system interrupt key (refer to the SAS[®] Companion for your system) and choosing to stop computations. For example, you can press Control-C on a Workstation or Control-Break on a PC to terminate exact computations.

Note: The network algorithm, as implemented in the NPAR1WAY procedure, represents the scores using their full precision. Although it is possible to increase the speed of the algorithm by rounding the scores to a small number of decimal digits, this is not done by the procedure since it can result in loss of accuracy in the p -values.

Summary

This paper introduced the new EXACT statement in the

NPAR1WAY procedure. This statement is available beginning with Release 6.11 of the SAS System, and in Release 6.09E of the SAS Systems on mainframes and minicomputers.

This statement may be used to compute p -values for two-sample tests using Wilcoxon-Mann-Whitney, Median, Van der Waerden and Savage scores, or for multi-sample tests using Kruskal-Wallis, Brown-Mood, k -sample Van der Waerden and k -sample Savage scores. Examples are provided to illustrate the importance of using exact, rather than asymptotic, p -values.

Acknowledgements

We are grateful to Robert N. Rodriguez and Donna M. Sawyer of the Applications Division at SAS Institute for their valuable assistance in the preparation of this manuscript.

SAS and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

References

- Agresti, A. (1992), "A Survey of Exact Inference for Contingency Tables," *Statistical Science*, 7(1), 131-177.
- Agresti, A., Mehta, C.R. and Patel, N.R. (1990), "Exact Inference for Contingency Tables with Ordered Categories," *Journal of American Statistical Association*, 85, 453-458.
- Fisher, R.A. (1934), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.
- Hajek, J. (1969), *A Course In Nonparametric Statistics*, San Francisco: Holden-Day.
- Hirji, K.F., Mehta, C.R. and Patel, N.R. (1987), "Computing Distributions for Exact Logistic Regression," *Journal of American Statistical Association*, 82, 1110-1117.
- Hollander, M. and Wolfe D.A. (1973), *Nonparametric Statistical Methods*, New York: John Wiley & Sons.
- Mehta, C.R. and Patel, N.R. (1983), "A Network Algorithm for Performing Fisher's Exact Test in rxc Contingency Tables," *Journal of American Statistical Association*, 78, 427-434.
- Mehta, C.R. and Patel, N.R. (1986), "FEXACT: A Fortran Subroutine for Fisher's Exact Test in Unordered rxc Contingency Tables," *ACM Transactions on Mathematical Software*, 12, 154-161.
- Mehta, C.R., Patel, N.R. and Gray, R. (1985), "Computing an Exact Confidence Interval for the Common Odds Ratio in Several 2 by 2 Contingency Tables," *Journal of American Statistical Association*, 80, 969-973.
- Mehta, C.R., Patel, N.R. and Senchaudhuri, P. (1991), "Exact Stratified Linear Rank Tests for Binary Data," *Computing Science and Statistics: Proceedings of the 23rd symposium on the Interface*, ed. E.M. Keramidas, 200-207.
- Mehta, C.R., Patel, N.R. and Tsiatis, A.A. (1984), "Exact Significance Testing to Establish Treatment Equivalence with Ordered Categorical Data," *Biometrics*, 40, 819-825.
- SAS Institute Inc., (1990). *SAS/STAT User's Guide, Volume 2, Version 6, Fourth edition*, Cary, NC: SAS Institute Inc.
- Valz, P.D. and Thompson, M.E. (1994), "Exact Inference for Kendall's S and Spearman's Rho with Extensions to Fisher's Exact Test in rxc Contingency Tables," *Journal of Computational and Graphical Statistics*, 3(4), 459-472.

Authors

Anupama Narayanan, SAS Institute Inc., SAS Campus Drive, Cary, NC, 27513. Phone (919)677-8000 ext 5382. FAX (919)677-4444. Email sasanu@unx.sas.com

Donna L. Watts, SAS Institute Inc., Atlanta Plaza, Suite 3390, 950 E. Paces Ferry Rd N.E., Atlanta, GA 30326. Phone (404)814-2560 ext 238. FAX (919)814-2566. Email sasdlw@unx.sas.com