Behavior Research Methods, Instruments, & Computers
1996, 28 (2), 315-318

Implementation of nonparametric multivariate statistics with S

CHING-FAN SHEU and SUZANNE O'CURRY
DePaul University, Chicago, Illinois

Multivariate nonparametric statistical methods have not been widely used by psychologists. One reason for this may have been that the usual general-purpose packages do not provide easy implementation of these methods. In this article, we briefly describe the multivariate extensions of the sign, signed-rank, and rank-sum tests and use S, a programming environment for data analysis, to implement these statistical procedures. Three numerical examples are used to illustrate the flexibility and efficiency of these computations in S.

Two observations motivate us to write this article: (1) nonparametric multivariate methods are relatively unknown to psychologists, and (2) most statistical software packages do not provide easy implementation of these methods. There is, most likely, a relation between the two.

A main feature of nonparametric methods is the weak set of assumptions required for their validity. This is particularly important to social scientists, for often the measurements used in our research do not go beyond assigning ranks to observations. Sometimes, we may only be able to note the signs of difference between treatments.

Siegel's (1956) and Siegel and Castellan's (1988) *Nonparametric Statistics for the Behavioral Sciences* are widely known to psychologists. However, they did not discuss multivariate methods. On the other hand, the standard textbooks on multivariate statistics do not treat nonparametric methods (e.g., Morrison's, 1976). Thus, few psychologists have had the opportunity to become familiar with the multivariate extension of sign test, Wilcoxon signed rank test, and Mann–Whitney–Wilcoxon test.

In this article, we will briefly describe these useful nonparametric tests and implement them in a highly flexible programming environment called S (Becker, Chambers, & Wilks, 1988). We will illustrate the numerical computation with small data sets.

Presently, many statistical software packages meet most of the needs of a psychologist in his/her data analysis. The best known of these are, perhaps, Statistical Package for the Social Sciences (SPSS), Statistical Analysis System (SAS), Biomedical Computer Programs (BMDP), and MINITAB. Each of these packages offers a wide range of statistical techniques and has extensive user manuals. One or another of the software packages is often used in advanced undergraduate or first-year graduate level statistics courses for psychologists. Because these software packages are designed to allow for easy applications of the standard methods (see, e.g., Hays, 1991), they are less amenable to users who wish to develop individual programs for specialized purposes.

To perform multivariate tests, a software package must provide flexible ways to represent and manipulate matrices. To facilitate matrix calculations, SPSS has a matrix macro facility (Norusis, 1993) and SAS has SAS/IML software (SAS Institute, 1990). However, the two packages are no longer user-friendly in these aspects. The users are usually restricted to the use of standard multivariate methods. A user wishing to implement his/her own statistical procedures might find the task easier to achieve in some general-purpose mathematical software packages, such as Mathematica, Matlab, and Maple. Each of the packages provides an interactive environment in which calculations involving matrices can be done efficiently. The drawback is that they are not dedicated statistical packages, and users must write their own basic statistical routines. These considerations lead us to choose S for the implementation of the multivariate nonparametric tests.

S is an interactive programming environment for data analysis and graphics. Users can either use its built-in routines to run traditional univariate tests such as linear regression and analysis of variance or they can program it to perform user-specified computation. Though we do not discuss its extensive graphical facilities in this paper, S also offers many flexible ways to explore data through visualization. Our purpose is to show how S makes learning and doing multivariate nonparametric tests simple and easy.

Our discussion of the multivariate extensions of the sign, signed-rank, and rank-sum tests will follow the accounts available in Hettmansperger (1984) and Leach (1991). The S codes for these statistical tests are presented in three separate listings. They are written to mirror as closely as possible the computational steps presented in the text. Obviously, they can be much improved in both efficiency and elegance. The reader need only look up the keywords in Becker et al. (1988) to follow our examples. Everitt

This work is partially supported by a competitive instructional grant from the Quality of Instruction Council of DePaul University to C.-F.S. We thank two anonymous reviewers for comments on a previous draft of this paper. Correspondence concerning this paper should be addressed to C.-F. Sheu, Department of Psychology, DePaul University, 2219 N. Kenmore Ave., Chicago, IL 60614-3504 (e-mail: csheu@condor.depaul.edu).

(1994) gives a brief introductory guide to S (or S-Plus). Detailed discussion can be found in Venables and Smith (1992) and in Spector (1994).

MULTIVARIATE SIGN TEST

To illustrate the multivariate sign test, we use scores of 10 women on the Beck Depression Inventory (BDI) and Self-Esteem Scale collected by Ainscough and Toon at the beginning and end of a group for adult survivors of child sexual abuse. This data set is reported in Leach (1991) and is reproduced in part in Table 1.

The data array consists of a random sample of 10 observations on two variables. Each variable is measured twice: before and after group treatment. If a person improves, a decrease on his/her BDI score and/or an increase in self-esteem score should be observed. In other words, this is a standard paired comparison design. Subject 1's BDI score is an outlier and suggests that a nonparametric procedure is more appropriate than a parametric one.

A sign statistic usually involves counting up the number of positive values, $N+$, in a sample. An alternative statistic, D, is the difference between the numbers of positive signs and negative signs. It is easy to see that $D = 2N^+ - n$, where n is the number of observations in the sample. Although the two forms are equivalent mathematically, the latter is more convenient in the multivariate case. For data in Table 1, the observed difference between the number of positive and negative difference scores for the BDI is 6 and that for self-esteem is 8. In other words, we compute a univariate statistic on each of the differences to obtain the vector of observed statistics,

$$D = \begin{bmatrix} -6 \\ 8 \end{bmatrix}.$$

In general, however, more than two variables may be present in a problem, then $D' = [D_1, D_2, \ldots, D_p]$ will be a random vector of size $p \times 1$, where p is the number of variables. The null hypothesis is that the population median scores for each difference measurement before and after the group are all zero. The limiting distribution for $D/\sqrt{n}$ under the null hypothesis is a multivariate normal distribution with zero mean vector and a covariance matrix V. The elements of V are estimated by

$$\hat{v}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \mathrm{sgn}(X_{ik})\mathrm{sgn}(X_{jk}),$$

where $\mathrm{sgn}(X_{ik})$ is the sign of the difference score for the kth subject on the ith variable. For the example, the estimated covariance matrix is of the following form:

$$\hat{V} = \frac{1}{10}\begin{bmatrix} \mathrm{sgn}(X_{1,1}) & \mathrm{sgn}(X_{1,2}) & \ldots & \mathrm{sgn}(X_{1,10}) \\ \mathrm{sgn}(X_{2,1}) & \mathrm{sgn}(X_{2,2}) & \ldots & \mathrm{sgn}(X_{2,10}) \end{bmatrix} \begin{bmatrix} \mathrm{sgn}(X_{1,1}) & \mathrm{sgn}(X_{2,1}) \\ \mathrm{sgn}(X_{1,2}) & \mathrm{sgn}(X_{2,2}) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathrm{sgn}(X_{1,10}) & \mathrm{sgn}(X_{2,10}) \end{bmatrix}.$$

Then, under the null hypothesis, the statistic

$$D^* = D^t(n\hat{V})^{-1}D$$

is asymptotically a chi-square distribution with degrees of freedom equal to the number of variables.

Leach (1991) gave a value of 6.44 for D^* in this example. Listing 1 implements the computation in S. The resulting chi-square approximation yields the same value (within rounding error) with a corresponding p value of 3.99%.

Table 1
Before and After Scores on the Beck Depression Inventory and Self-Esteem Scale for 10 Participants

	Beck Depression Inventory		Self-Esteem Scale	
Subject	Before	After	Before	After
1	24	63	20	17
2	2	3	61	62
3	6	3	30	54
4	15	11	35	41
5	6	0	34	63
6	10	0	48	63
7	32	18	17	24
8	23	1	31	41
9	36	8	22	41
10	31	2	22	44

Note—Unpublished data from Ainscough and Toon; see Leach (1991).

MULTIVARIATE ONE-SAMPLE SIGNED-RANK TEST

In the example above, instead of simply noting the signs, we might also take into account the magnitude of departures from the median specified under the hypothesis of no difference. This requires us to assume that the underlying population distribution is symmetric. We then consider a form of the Wilcoxon signed rank statistic,

$$T_i = \sum_{k=1}^{n} \frac{R_{ik}\mathrm{sgn}(X_{ik})}{n+1},$$

where the numerator is the rank of the absolute value of the kth difference score on the ith variable with sign attached, and n is the number of difference scores. For data in Table 1, the statistic above for the BDI score is -3 and that for self-esteem score is 4.63, yielding a vector $T^t = [-3 \quad 4.63]$. For a large sample, the statistic $T/\sqrt{n}$ is approximately a multivariate normal distribution with zero mean vector and a covariance matrix V. The entries of V are estimated by

$$\hat{v}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \frac{R_{ik}\mathrm{sgn}(X_{ik})}{n+1}\frac{R_{jk}\mathrm{sgn}(X_{jk})}{n+1}.$$

Under the null hypothesis, the quadratic form statistic, $T^* = T^t(n\hat{V})^{-1}T$, is asymptotically distributed as a chi-square distribution with degrees of freedom equal to the number of variables. Listing 2 shows the implementation

Listing 1
An S Session to Compute Multivariate Sign Test Using Data of Table 1

```
> bdi ← matrix(scan("bdi.scores"), ncol=2, byrow=T)
> estm ← matrix(scan("esteem.scores"), ncol=2, byrow=T)
> # We assume the BDI and Self-Esteem scores are stored in two separate files under the
  current working directory. bdi and estm are 10 by 2 matrices of scores
> dfBdi ← bdi[,2] - bdi[,1]
> dfEstm ← estm[,2] - estm[,1]
> # This generates a vector of difference scores for BDI and Self-Esteem scores
> sign ← function(M) ifelse(M < 0, -1, 1)
> # This creates a user-defined sign function
> sgnDf ← cbind(sign(dfBdi), sign(dfEstm))
> # The two vectors of signs are bound together column-wise
> dfSgn ← cbind(sum(sgnDf[,1]), sum(sgnDf[,2]))
> # This is a row vector of difference between the number of positive and negative signs
> dfSgnV ← crossprod(sgnDf)
> # This is nV̂
> starD ← dfSgn %*% solve(dfSgnV) %*% t(dfSgn)
> # This calculates the test statistic D*. In S, solve inverts a matrix. Note that dfSgn is a
  row vector
> 1 - pchisq(starD, df=2)
       [,1]
[1] 0.03986615
> # pchisq is the cumulative chi-square distribution function
```

Note—The S prompt is " >". Lines preceded by # are comments.

of the calculations in S. For data in Table 1, we obtain a $T^* = 7.10$ corresponding to a p value of 2.87%, which is in close agreement with the result of the sign test.

TWO-SAMPLE MULTIVARIATE RANK-SUM TEST

Morrison (1976, p. 167) reported a drug trial in which 10 mice were assigned at random to a control group and 12 to a treatment group. The levels of two biochemical compounds found in the brains of mice were recorded in Table 2. We use this example to illustrate a multivariate version of the Mann–Whitney–Wilcoxon test. The null hypothesis is that the treatment had no effect on the levels of biochemical compounds. The assumption is that the two samples come from the same population distribution. The alternative hypothesis is that the samples are from populations differing only in a shift of location (i.e., mean or median).

The rank-sum approach is to make a joint ranking of observations from the two samples and sum the ranks associated with one sample, usually the smaller of the two. For the example, we compute the ranks of the first compound of the pooled sample (control and treatment combined). The ranks for the second compound are similarly obtained. Ties are assigned average rank. In general, let $N = n + m$ be the size of two groups combined, and let n be the size of the smaller group. Hettmansperger (1984) uses the centered rank sum statistic

$$U_i = \sum_{k=1}^{n}\left[\frac{R_{ik}}{N+1} - \frac{1}{2}\right],$$

where R_{ik} is the rank of the smaller sample observations in the combined sample of the ith compound. For the data in Table 2, $U^t = [-1.80, \quad 1.43]$. The asymptotic distribution of $U/\sqrt{N}$ under the null hypothesis is a multivariate normal with zero mean vector and a covariance matrix V. The elements of the covariance matrix V are estimated by

$$\hat{v}_{ij} = \frac{mn}{N^2(N-1)(N+1)^2}\left\{\sum_{k=1}^{N} R_{ik}R_{jk} - \frac{N(N+1)^2}{4}\right\},$$

where $\sum_{k=1}^{N} R_{ik}R_{jk}$ is the sum of the cross products of the rankings in the pooled sample. The test statistic $U^* = U^t(N\hat{V})^{-1}U$ is asymptotically chi-square under the null hypothesis with degrees of freedom equal to the number

Listing 2
A Continuing S Session to Compute Multivariate Sign-Rank Test Using Data of Table 1

```
> rnkDfBdi ← rank(abs(dfBdi))
> rnkDfEstm ← rank(abs(dfEstm))
> # This ranks the difference scores in terms of absolute values
> rnkDf ← cbind(rnkDfBdi, rnkDfEstm)
> sgnDf ← cbind(sign(dfBdi), sign(dfEstm))
> # Generates a matrix of 10 × 2 ranks and a corresponding matrix
  of signs
> sgnRnkDf ← rnkDf * sgnDf
> # Attaches signs to corresponding ranks
> nObs ← nrow(bdi)
> # Defines the size of the sample
> sgnRnk ← sgnRnkDf/(nObs+1)
> sgnRnkT ← cbind(sum(sgnRnk[,1]), sum(sgnRnk[,2]))
> # The sign-rank statistic as a row vector
> cvT ← crossprod(sgnRnk)
> # This is nV̂
> starT ← sgnRnkT %*% solve(cvT) %*% t(sgnRnkT)
> # This is the test statistic
> 1-pchisq(starT, df=2)
       [,1]
[1,] 0.02867436
> # The corresponding P-value using chi-square approximation
```

Note—The S prompt is " >". Lines preceded by # are comments.

Table 2
Levels of Biochemical Compound in Micrograms per Gram of Brain Tissue

Control Group		Treatment Group	
Compound 1	Compound 2	Compound 1	Compound 2
1.21	0.61	1.40	0.50
0.92	0.43	1.71	0.39
0.80	0.35	1.23	0.44
0.85	0.48	1.19	0.37
0.98	0.42	1.38	0.42
1.15	0.52	1.71	0.45
1.10	0.50	1.31	0.41
1.02	0.53	1.30	0.47
1.18	0.45	1.22	0.29
1.09	0.40	1.00	0.30
		1.12	0.27
		1.09	0.35

Listing 3
An S Session to Compute Multivariate Rank-Sum Test Using Data from Table 2

```
> cntrl ← matrix(scan("mice.cntrl.dat"), ncol=2, byrow=T)
> trtmnt ← matrix(scan("mice.trtmnt.dat"), ncol=2, byrow=T)
> # Reads data in matrix form
> nC ← 10; nT ← 12; N ← nC+nT
> # Sets number of observations in each group and total number
  of observations
> rnkC1T1 ← rank(append(cntrl[,1], trtmnt[,1]))
> rnkC2T2 ← rank(append(cntrl[,2], trtmnt[,2]))
> # Assigns ranks to observations of each compound in the pooled
  sample
> rnkC1 ← rnkC1T1[1:nC]
> rnkC2 ← rnkC2T2[1:nC]
> # Picks out the ranks of the control group
> sumRnkC1 ← sum((rnkC1/(N+1))-0.5)
> sumRnkC2 ← sum((rnkC2/(N+1))-0.5)
> U ← cbind(sumRnkC1, sumRnkC2)
> # This is the rank-sum statistic as a row vector
> rnkCT ← cbind(rnkC1T1, rnkC2T2)
> # A 22 by 2 matrix of joint rankings
> coefM ← nC*nT/((N+1)*(N+1)*(N-1)*N*N)
> coefL ← (N*(N+1)*(N+1))/4
> # Defines coefficients in covariance formula
> cvU ← coefM*((crossprod(rnkCT))-coefL)
> # This is V̂.
> starU ← U %*% solve(N*cvU) %*% t(U)
> # This is the test statistic
> 1-pchisq(starU, df=2)
          [,1]
[1,] 0.0007717609
```

Note—The S prompt is " >". Lines preceded by # are comments.

of variables. For the example, the observed value of U^* is about 14.33 corresponding to a p value of less than 0.1%. Listing 3 implements the above calculations in S.

CONCLUDING REMARKS

This article presents multivariate extensions of the sign, sign-rank, and rank-sum tests and their implementations in S. Because the tests are based on very similar mathematical arguments, and because the S language has a set of powerful operators and functions, none of the code that we developed in the listings has more than 20 lines. We believe that S is especially valuable in enabling students to understand the links between different tests and the underlying statistical procedures, because the software makes the computation transparent.

REFERENCES

BECKER, R. A., CHAMBERS, J. M., & WILKS, A. R. (1988). *The new S language: A programming environment for data analysis and graphics*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

EVERITT, B. S. (1994). *A handbook of statistical analysis using S-PLUS*. New York: Chapman & Hall.

HAYS, W. L. (1991). *Statistics* (5th ed.) Fort Worth, TX: Harcourt Brace.

HETTMANSPERGER, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.

LEACH, C. (1991). Nonparametric methods for complex data sets. In A. D. Lovie (P. Lovie, Ed.) *New developments in statistics for psychology and the social sciences* (Vol. 2, pp. 1-18). London: Routledge.

MORRISON, D. F. (1976). *Multivariate statistical methods*. New York: McGraw-Hill.

NORUSIS, M. J. (1993). *SPSS for UNIX advanced statistics* (Release 5.0) [computer software]. Chicago: SPSS Inc.

SAS INSTITUTE (1990). *SAS/IML user's guide* (Version 6 ed.). Cary, NC: Author.

SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

SIEGEL, S., & CASTELLAN, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

SPECTOR, P. (1994). *An introduction to S and S-PLUS*. Belmont, CA: Duxbury Press.

VENABLES, W. N., & SMITH, D. (1992). *Notes on S-PLUS: A programming environment for data analysis and graphics*. Unpublished manuscript. Available: venables@stats.adelaide.edu.au.

(Manuscript received November 13, 1995;
accepted for publication December 20, 1995.)