

PERMUTATION TESTS FOR CORRELATION
BETWEEN TWO DISTANCE MATRICES

E. Jacquelin Dietz

Department of Statistics, Box 5457, North Carolina State University,
Raleigh, North Carolina 27650, U.S.A.

SUMMARY

Biologists frequently summarize multivariate data from n populations by computing some measure of distance between populations. The problem then arises of comparing two such pairwise distance matrices based on different characters. A permutation test for correlation between distance matrices is proposed. This test, based on Kendall's tau statistic, is compared to Pearson's product-moment and Spearman's rho tests for the same problem. Examples are presented, and computation is discussed.

1. Introduction

Are geographically separated populations more different genetically than nearby populations? Do anthropometric and linguistic data reflect similar relationships among populations? In general, do relationships among populations found for one character "correspond" to those for another character? Questions of this type are familiar to biologists, and many authors have proposed methods for comparing two sets of multivariate observations. The usual approach reduces each set of multivariate

Key words: Permutation tests; Correlation; Distance matrices; Kendall's tau; Measures of association; Distribution free.

data to some measure of generalized distance between populations. The problem is then to compare the resulting pairwise distance matrices.

Some authors (Howells 1966, McKechnie, Ehrlich and White 1975) have simply calculated the sample correlation coefficient r between the pairwise distances. The normality assumption required in order to test the significance of r is suspect for many distance measures (see e.g., Friedlaender *et al.* 1971, p. 267). More problematic, however, are the dependencies among the $\binom{n}{2}$ pairwise distances for n populations. The degrees of freedom for a correlation test are surely overestimated when the test is applied to such data. These difficulties have been discussed (Friedlaender *et al.* 1971, Spielman 1973), but valid tests for correlation between distance measures are not widely known.

Alternative techniques for comparing sets of data have been proposed. Spielman (1973) uses networks or trees derived from distance measures to represent relationships among populations. Using path lengths of the networks, he tests whether two sets of data give rise to similar cluster structures. Corruccini, Cronin and Ciochon (1979) perform a principal coordinates analysis for each set of distance data. Two sets are then compared by correlating the positions of the populations on the most significant principal coordinate axes. Spielman (1973) also adapts the method of Schönemann and Carroll (1970) to test a different kind of correspondence between two sets of data. This test of geometric congruence fits one matrix of multivariate observations to another by linear transformations. Spielman's techniques are difficult to apply in practice and require complex computer simulations to generate needed null distributions. None of the analyses mentioned here are direct or intuitive;

interpretation of results seems difficult for all of them. In fact, Spielman's two methods can lead to very different conclusions for the same data.

We propose a new test for correlation between two distance matrices. This permutation test requires no assumptions about the distributions of the distance measures and is invariant under monotone transformations of either measure. We address the intuitively appealing question: If two populations are more "distant" than another two populations, according to some trait, will the same relationship be likely to hold for a different trait? A test proposed by Mantel (1967) and Mantel and Valand (1970) for a similar problem is described and compared to the new test.

2. Procedure

2.1 Test Statistic

Let n be the number of populations, and let X_{ij} and Y_{ij} , $1 \leq i \neq j \leq n$, be the distances between populations i and j for two distance measures. The only restriction placed on these measures is $X_{ij} = X_{ji}$ and $Y_{ij} = Y_{ji}$ for all i and j . There are $\binom{n}{2}$ distinct (X,Y) pairs.

Define

$$\tau = P\{(X_{ij} - X_{kl})(Y_{ij} - Y_{kl}) > 0\} - P\{(X_{ij} - X_{kl})(Y_{ij} - Y_{kl}) < 0\} .$$

We test H_0 , the hypothesis that X and Y are independent, versus the alternative $\tau > 0$. Alternatives $\tau < 0$ and $\tau \neq 0$ are handled similarly.

Letting

$$\text{sign}(a) = \begin{cases} +1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases},$$

define the test statistic

$$K = \sum_{i,j,k,l} \text{sign}(X_{ij} - X_{kl})(Y_{ij} - Y_{kl}),$$

where summation is over the $\binom{n}{2}$ distinct pairs of distances. Note that K is Kendall's (1975) statistic S , applied to the distances.

Ordinarily, Kendall's procedure is applied to independent bivariate observations $(U_1, V_1), \dots, (U_m, V_m)$ to test independence between U and V . Then

$$S = \sum_{i=1}^{m-1} \sum_{j=i+1}^m \text{sign}(U_i - U_j)(V_i - V_j).$$

When U and V are independent, given the U_i , the $m!$ permutations of the V_i are equally likely. S depends only on the ranks of the U_i among themselves and the V_i among themselves, and (assuming no ties) the test based on S is distribution-free.

In our situation, the $\binom{n}{2}$ (X, Y) pairs are not mutually independent, and the $\binom{n}{2}!$ possible permutations of the Y 's are not all equiprobable under H_0 . In fact, only $n!$ of the possible permutations are equally likely.

2.2 Permutation Distribution

Let $(P_{(1)}, \dots, P_{(n)})$ be a permutation of the integers $(1, \dots, n)$. Let Π be the set of all $n!$ such permutations. Under H_0 , given the X_{ij} , there are $n!$ equiprobable permutations of the Y_{ij} 's, each corresponding to a member of Π . For each $P \in \Pi$, pair X_{ij} with $Y_{P(i) P(j)}$, for every i and j . These permutations of the Y 's correspond to relabeling the n populations for the Y distance measure.

Under H_0 , the $n!$ corresponding values of K each have probability $1/n!$. Although K depends only upon the ranks of the X 's among themselves and the Y 's among themselves, its distribution is conditional upon the (ranked) distance matrices for X and Y . The test based on K is thus a conditional test. Unconditionally, under H_0 ,

$$\begin{aligned} E(K) &= \frac{1}{n!} \sum_{P \in \Pi} \sum_{i,j,k,\ell} \text{sign}(X_{ij} - X_{k\ell}) (Y_{P(i) P(j)} - Y_{P(k) P(\ell)}) \\ &= \frac{1}{n!} \sum_{i,j,k,\ell} \text{sign}(X_{ij} - X_{k\ell}) \sum_{P \in \Pi} \text{sign}(Y_{P(i) P(j)} - Y_{P(k) P(\ell)}) = 0 . \end{aligned}$$

Notice that the distribution of K does not depend upon which measure is called X or Y . Also, tied observations present no difficulty.

Since large values of K correspond to large values of τ , the significance level α of an observed statistic is the proportion of permutations yielding a value of K greater than or equal to the observed value. For $n \leq 8$, we can find α for an observed K exactly. For $n \geq 9$, $n!$ is very large, and complete enumeration of the permutations is not feasible. Then random sampling of the $n!$ permutations allows us to estimate α .

2.3 Large Sample Test

The idea of sampling a subset of permutations in a permutation test appears frequently in the literature; see, for instance, Dwass (1957), Chung and Fraser (1958) and Boyett and Shuster (1977). For $n \geq 9$, we estimate the significance level α of an observed statistic using a random sample of m of the $n!$ permutations, chosen with replacement. The sample estimate is unbiased with variance $\alpha(1-\alpha)/m$.

Another criterion for choosing m is provided by Dwass (1957). His results for the two sample location problem apply to our situation without modification. He shows that if both tests use a significance level of .05, a test based on 1000 permutations is 94.5% as powerful as the test based on full enumeration. The corresponding comparison for $m = 5000$ is 97.5%.

3. Comparison with Test of Mantel and Valand

Mantel (1967) proposed a permutation test for detecting clustering of disease cases in time and space. The same procedure was discussed in a more general setting by Mantel and Valand (1970). Letting X_{ij} and Y_{ij} be spatial and temporal measures, respectively, between points i and j , the proposed test statistic is $Z = \sum_{i \neq j} X_{ij} Y_{ij}$. It is not required that $X_{ij} = X_{ji}$ and $Y_{ij} = Y_{ji}$. (This restriction could easily be removed in the new test, also. The reduction in computing time obtained by the symmetry assumption justifies its use, since in most applications symmetry is present.) The hypothesis of no clustering corresponds to the hypothesis that the locations in space and locations in time are randomly paired. Thus the null distribution of Mantel's statistic is generated by the same $n!$ permutations of the observations proposed in Section 2.2.

To test the significance of an observed statistic, Mantel assumes that Z is approximately normally distributed. He gives formulas for the permutational mean and variance of Z and suggests that Z be standardized and referred to tables of the normal distribution. Mielke (1978) shows that Z is not, in general, asymptotically normal and gives a beta approximation suitable for certain special cases.

Such distributional assumptions can be avoided entirely by enumerating the permutation distribution of Z , using sampling of permutations for large samples. We then have the following family of permutation tests for correlation. Mantel's test represents a Pearson product-moment correlation approach. Mantel's test, when applied to the within-sample ranks of the X 's and Y 's, corresponds to a Spearman's rho approach; the new test is a Kendall's τ approach.

The Pearson statistic Z is highly dependent upon the specific distance measures X and Y . Mantel (1967) recommends that a reciprocal transformation be applied to both temporal and spatial distances. His examples show that this order-preserving transformation can dramatically affect the outcome of his test. For this reason, the Spearman and Kendall-type statistics seem preferable for comparing distance measures whose actual magnitudes may be arbitrary.

The Kendall statistic K is easily interpreted in terms of a parameter τ , a characteristic not shared by the Spearman-type statistic. Also, K can be calculated from data which cannot be ranked. Imagine a situation where pairwise comparisons of distances are the only data available. For instance, we may know that populations i and j are more similar according to some character than are populations k and l , etc. Even with inconsistencies among these pairwise comparisons, the K

statistic can be computed and tested for significance. See Kendall (1975, Chapter 11) for a discussion of such paired comparisons data.

4. Examples

Some properties of the three permutation tests for correlation can be illustrated using data from Spielman (1973). He presents anthropometric, genetic, and geographic distance data for 19 villages of Yanomama Indians.

Many authors have applied conventional correlation tests to distance data. The misleading nature of such tests can be demonstrated using Spielman's marker gene and anthropometric data. Significance levels for the three permutation tests were estimated using 2000 random permutations. Conventional Kendall, Spearman, and Pearson correlation tests were performed and tested for significance, as though the $\binom{19}{2} = 171$ pairs of distances were independent. The observed statistics for these latter tests were $K = 1788$, $r_s = .191$ and $r = .113$, respectively. The estimated levels for the permutation tests and the apparent levels for the conventional tests were as follows:

Test	Estimated level (Permutation test)	Apparent level (Conventional test)
Kendall	0.0765	0.0084
Spearman	0.0645	0.0064
Pearson	0.2010	0.0704

As expected, the significance of the observed statistics is overestimated by the conventional test procedures. Notice also the discrepancy in significance between the Pearson statistic and the two rank statistics.

When $n = 8$, both small sample and large sample permutation tests are applicable. These options were compared for Spielman's anthropometric and SFA data, using the first eight villages. (The SFA, "Serological for Anthropometrics", data are genetic distances based on those individuals for whom anthropometric measurements were also obtained - a subset of the total sample.) Enumeration of all $8! = 40320$ permutations provided exact significance levels for the three tests. Estimates of these levels based on 5000 random permutations were also obtained, with the following results:

Test	Exact level	Estimated level
Kendall	0.0047	0.0034
Spearman	0.0060	0.0058
Pearson	0.0032	0.0024

For all three tests, the estimated levels are well within two standard deviations of the true values.

5. Computation

A Fortran program for performing the Pearson, Spearman, and Kendall correlation tests is available upon request. The program finds the exact significance level of each observed statistic for small samples and estimates the levels by sampling for large samples.

For small samples, the permutations of $(1, \dots, n)$ are obtained using an algorithm of Boothroyd (1967). For large samples, random permutations are generated using a shuffling algorithm described in Knuth (1969, p. 125). The uniform random numbers required by that algorithm are generated by a multiplicative congruential method (Schrage 1979).

Since the feasibility of a permutation test depends upon the computing time required, the program was run for several sets of test data on the IBM 370/165 at the Triangle Universities Computation Center. Calculation of exact significance levels for an example with $n = 7$ required 31.5 seconds of central processing units (cpu) time. The example in Section 4 with $n = 8$ required 38.5 seconds of cpu time for 5000 random permutations and 2:22.4 minutes of cpu time for exact solutions. An example with $n = 12$ and $m = 5000$ required 1:37.9 minutes; the example in Section 4 with $n = 19$ and $m = 2000$ required 3:35.6 minutes.

For each permutation, the Kendall statistic is the sum of $\binom{n}{2}$ terms; the Spearman and Pearson statistics are each (assuming $X_{ij} = X_{ji}$ and $Y_{ij} = Y_{ji}$) the sum of $\binom{n}{2}$ terms. Thus, the Kendall test requires more computing time than the other tests. Since for large samples the Kendall and Spearman tests seem to give similar results, the Spearman test may be the preferred procedure when n is large.

ACKNOWLEDGEMENTS

I wish to thank Richard W. Morris for suggesting this problem and for many helpful discussions. John F. Monahan provided valuable assistance with computing algorithms.

REFERENCES

- Boothroyd, J. (1967). Algorithm 29, Permutation of the elements of a vector. *The Computer Journal* 60, 311.
- Boyett, J. M. and Shuster, J. J. (1977). Nonparametric one-sided tests in multivariate analysis with medical applications. *Journal of the American Statistical Association* 72, 665-668.
- Chung, J. H. and Fraser, D. A. S. (1958). Randomization tests for a multivariate two-sample problem. *Journal of the American Statistical Association* 53, 729-735.
- Corruccini, R. S., Cronin, J. E. and Ciochon, R. L. (1979). Scaling analysis and congruence among anthropoid primate macromolecules. *Human Biology* 51, 167-185.
- Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics* 28, 181-187.
- Friedlaender, J. S., Sgaramella-Zonta, L., Kidd, K. K., Lai, L. Y. C., Clark, P. and Walsh, R. J. (1971). Biological divergences in South-Central Bougainville: An analysis of blood polymorphism gene frequencies and anthropometric measurements utilizing tree models, and a comparison of these variables with linguistic, geographic, and migrational "distances." *American Journal of Human Genetics* 23, 253-270.
- Howells, W. W. (1966). Population distances: Biological, linguistic, geographical, and environmental. *Current Anthropology* 7, 531-540.
- Kendall, M. (1975). *Rank Correlation Methods*. Charles Griffin and Company, Ltd., London.
- Knuth, D. E. (1969). *The Art of Computer Programming, Vol. 2*. Addison-Wesley Publishing Co., Inc., Massachusetts.

- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27(I), 209-220.
- Mantel, N. and Valand, R. S. (1970). A technique of nonparametric multivariate analysis. *Biometrics* 26, 547-558.
- McKechnie, S. W., Ehrlich, P. R. and White, R. R. (1975). Population genetics of *Euphydryas* butterflies. I. Genetic variation and the neutrality hypothesis. *Genetics* 81, 571-594.
- Mielke, P. W. (1978). Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique. *Biometrics* 34, 277-282.
- Schönemann, P. H. and Carroll, R. M. (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* 35, 245-255.
- Schrage, L. (1979). A more portable Fortran random number generator. *ACM Transactions on Mathematical Software* 5, 132-138.
- Spielman, R. S. (1973). Differences among Yanomama Indian villages: Do the patterns of allele frequencies, anthropometrics and map locations correspond? *American Journal of Physical Anthropology* 39, 461-480.