Quality & Quantity **37:** 99–110, 2003.

Research Note

An Approximate Test for Homogeneity of Correlated Correlation Coefficients

TRIVELLORE RAGHUNATHAN
University of Michigan

Abstract. This paper develops and evaluates an approximate procedure for testing homogeneity of an arbitrary subset of correlation coefficients among variables measured on the same set of individuals. The sample may have some missing data. The simple test statistic is a multiple of the variance of Fisher r-to-z transformed correlation coefficients relevant to the null hypothesis being tested and is referred to a chi-square distribution. The use of this test is illustrated through several examples. Given the approximate nature of the test statistics, the procedure was evaluated using a simulation study. The accuracy in terms of the nominal and the actual significance levels of this test for several null hypotheses of interest were evaluated.

Key words: chisquare test, Fisher r-to-z, missing data

1. Introduction

A common situation in social science research involves a comparison a set of correlation coefficients between variables measured on the same subjects. For example:

(1) In an evaluation of several instruments for scoring a certain attribute, testing for the homogeneity of correlation coefficients between scores obtained using these instruments.
(2) Suppose that there are several, possibly nested instruments of differing length (hence, differing costs) for scoring current health status and the objective is to relate the current health status score to the medical cost or utilization. If all the current health status scores from different instruments are equally correlated to the dependent variable (medical costs) then the shortest instrument may be used to reduce costs.

An important difference between these two problems is that in the former, all possible $p(p-1)/2$ pairwise correlation coefficients among, say p, score variables are being tested for equality whereas in the second example only a subset of $p-1$ of the $p(p-1)/2$ possible pairwise correlation coefficients are involved in the null hypothesis. The means and variances are the nuisance parameters in the first problem and the means, variances and the remaining $(p-1)(p-3)/2$ correlation coefficients are the nuisance parameters in the second problem.

There are other examples in longitudinal studies where there is a need for comparing correlated correlation coefficients:

(3) Equality of the correlation coefficients between the same variable measured at several, say p, occasions.

(4) Equality of the correlation coefficients between two variables, say X and Y, measured at p occasions.

Again, these two examples differ because in the first, all possible pairwise correlation coefficients are involved in the null hypothesis, whereas in the second example, only a subset of correlation coefficients are involved in the null hypothesis.

The objective of this paper is to develop and evaluate a simple procedure for testing the equality among a set of correlated correlation coefficients that is applicable to both situations. These procedures also allow for missing data assumed to be generated by an ignorable missing data mechanism (Rubin, 1976). The proposed test statistic can be calculated using a hand-held calculator or a spreadsheet. In Section 2, we describe the test for the equality of all $p(1-p)/2$ pairwise correlation coefficients and illustrate its application using examples. In Section 3, the same test statistic is used illustrate the test for equality of arbitrary set of correlation coefficients. Given the approximate nature of the test, a simulation study was also conducted to study the actual significance levels of nominal 5% tests. The results from this simulation study are reported in Section 4 and finally, Section 5 concludes the paper with a technical justification for this procedure.

Many authors beginning with Pearson and Filon (1898) and Hotelling (1940) have considered the problem of comparing correlated coefficients. These authors focussed on four-variate and trivariate normal distributions respectively. Hotelling (1940) used restrictive assumptions on the covariance structure which was relaxed by Williams (1959). Dunn and Clark (1969, 1971) proposed tests, again for four-variate and trivariate normal distributions, based on Fisher's r-to-z transformation. They and others (Neill and Dunn (1975) and Steiger (1980)) also demonstrated the superiority of using z over r for small sample sizes and extreme sample correlations through simulations. Nonparametric procedures have been derived by Bennett (1978) and Choi (1977). Also, Steiger (1980) provides extensions and a comprehensive review of the literature.

The basic asymptotic results on the covariance between the correlation coefficients used in this article are derived in Olkin and Siotani (1976). Olkin and Finn (1990, 1995) discuss a variety of examples for testing the equality of special sets of correlated coefficients and develop appropriate likelihood ratio tests. These tests are computationally intensive and require special software packages to implement them, especially, when some data are missing. The approximate tests developed in this paper can be applied to those situations but in simple and easy-to-use form. Meng et al. (1992) developed an approximate test for the equality of correlation coefficient between a set of predictor variables and a common dependent variable

(Example 2 above). Also, Raghunathan et al. (1996) compare approximate procedures for comparing correlated correlation coefficients between two different pairs of variables measured on the same subjects (Example 4 above). The test statistic discussed in this paper can be applied to these situations as well.

2. Test for Equality of all Pairwise Correlation Coefficients

Based on a sample of size n (with, possibly, missing data on some variables), suppose that the variables X_i and X_j are observed on n_{ij} individuals and the sample correlation coefficient between these two variables is r_{ij} where $i \neq j = 1, 2, \ldots, p$. Suppose the population correlation coefficient between the variables X_i and X_j is ρ_{ij}.

Let $z_{ij} = 0.5 \times \log[(1+r_{ij})/(1-r_{ij})]$ denote Fisher's r-to-z transform of r_{ij}. Let $u_{ij} = \sqrt{n_{ij} - 3}(z_{ij} - \bar{z}_W)$ be the standardized Fisher transform of the correlation coefficient r_{ij} where $\bar{z}_W$ is the weighted average,

$$\bar{z}_W = \frac{\sum_{ij}(n_{ij} - 3)z_{ij}}{\sum_{ij}(n_{ij} - 3)}.$$

The proposed test statistic is

$$Q = \sum_{ij} u_{ij}^2. \tag{1}$$

The p-value for testing the hypothesis that all the correlation coeffcients are the same (i.e., H_o: $\rho_{ij} = \rho$; $i, j = 1, 2, \ldots, p, i \neq j$) is obtained by referring Q, to a chisquare distribution with ν degrees of freedom, where

$$\nu = \frac{p(p-1)}{2} - 1 - \frac{\bar{r}_W(p-2)(p\bar{r}_W + 2)}{(1+\bar{r}_W)^2},$$

where

$$\bar{r}_W = \frac{\exp(2\bar{z}_W) - 1}{\exp(2\bar{z}_W) + 1}.$$

We now illustrate the application of the above procedure using an example.

2.1. EXAMPLE 1

Table I gives the correlation matrix between scores on two tests on verbal (X_1, X_2), two tests on quantitative reasoning (X_3, X_4) and two tests on reading and comprehension (X_5, X_6) skills administered to 48 subjects. However, only 24 of the 48 subjects were given the second reading and comprehension test (X_6) because the remaining 24 subjects were asked some other questions in lieu of this test. The

Table I. Correlation matrix of test scores and the details for the calculations of test statistic and the degrees of freedom

Variable	Statictic	X_2	X_3	X_4	X_5	X_6
X_1	r	0.641	0.772	0.841	0.631	0.745
	z	0.758	1.020	1.221	0.743	0.962
	u	−0.925	0.834	2.181	−1.026	0.301
X_2	r		0.643	0.650	0.820	0.604
	z		0.759	0.775	1.157	0.699
	u		−0.925	−0.810	1.749.−0	901.
X_3	r			0.761	0.621	0.860
	z			0.996	0.727	1.292
	u			0.672	−1.136	1.821
X_4	r				0.627	0.742
	z				0.737	0.955
	y				−0.071	0.270
X_5	r					0.615
	z					0.717
	u					−0.821
$\sum(n-3)z$		34.111	80.055	134.640	151.377	97.146
$\sum(n-3)$		45	90	135	180	105
$\sum u^2$		0.856	1.551	5.865	6.549	4.966

objective is to test whether the correlation coefficients between scores measuring different functional abilities are equal. The sample size is $n_{ij} = 48$ everywhere except that it is 24 for the correlation coefficients in the last column. The number of correlation coefficients tested for homogeneity is $6 \times 5/2 = 15$.

Table I gives the intermediate results in the calculation of the test statistic Q and the degrees of freedom. First, Fisher's r-to-z transforms are computed which are given as the second entry in each cell of Table I. Based on the sums given in Table I,

$$\bar{z}_W = \frac{34.111 + 80.055 + 134.640 + 151.377 + 97.146}{45 + 90 + 135 + 180 + 225 + 105} = 0.896$$

and therefore,

$$\bar{r}_W = \frac{\exp(2 \times 0.896) - 1}{\exp(2 \times 0.896) + 1} = 0.714.$$

Given the value of z, n and $\bar{z}_W$, u is computed as $\sqrt{(n-3)}(z-\bar{z}_W)^2$ and given as the third entry in each cell of Table I. The value of Q-statistic given in Equation (1) is, therefore,

$$Q = 0.856 + 1.551 + 5.865 + 6.549 + 4.966 = 19.787,$$

and the degrees of freedom is

$$\nu = 15 - 1 - \frac{0.714 \times 4 \times (6 \times 0.714 + 2)}{1.714^2} = 7.888.$$

The resulting p-value is 0.0105 thus rejecting the null hypothesis of the homogeneity of the correlation coefficients between various test scores at the standard 0.05 level of significance.

3. Test for Equality of Arbitrary Set of Correlation Coefficients

We now consider the test for the equality of an arbitrary subset of correlation coefficients that involve nuisance correlation coefficients in the variance expression. Suppose that we are interested in testing the equality of only a subset of k of all possible $p(p-1)/2$ pairwise correlation coefficients. Let $d = k(k-1)/2$ be the number of pairs of correlation coefficients that are in the null hypothesis, of which, d_1 are between nonoverlapping pairs of variables and the remaining $d_2 = d - d_1$ pairs involve a common variable. For example, based on a sample from a $p = 4$ dimensional multivariate normal population, we are interested in testing the null hypothesis $H_o : \rho_{12} = \rho_{13} = \rho_{34}$. Here $k = 3$, $d = 3$ and the pairs (ρ_{12}, ρ_{13}), (ρ_{13}, ρ_{34}) are overlapping (that is, $d_2 = 2$) and the remaining pair (ρ_{12}, ρ_{34}) is nonoverlapping (that is, $d_1 = 1$). The correlation coefficients not involved in the hypothesis $(\rho_{14}, \rho_{23}, \rho_{24})$ are nuisance correlation coefficients, and these number $l = p(p-1)/2 - k$.

Let r_* be the median of the sample estimates of the l nuisance correlation coefficients. Let Q, $\bar{z}_W$ and $\bar{r}_W$ be exactly the same as defined before except that they are based only on the k sample correlation coefficients involved in the null hypothesis. The degrees of freedom for the chisquare statistic Q is

$$\nu = k - 1 - 2\frac{C_1 d_1 + C_2 d_2}{k},$$

where

$$C_1 = \frac{[\bar{r}_W^2 r_*^2 + (2r_* - \bar{r}_W^2)(1 - 2\bar{r}_W^2)]}{2(1 - \bar{r}_W^2)^2}$$

and

$$C_2 = \frac{2r_*^2}{(1 + \bar{r}_W)^2}.$$

Again, we illustrate the application of this procedure using two examples.

3.1. EXAMPLE 2

It has been argued in the psychometric literature that the speed of response-choice reaction time is positively associated with intelligence as measured by psychometric tests. Several studies have tried to relate several components of reaction times to intelligence. The correlation matrix given in Table III comes from one such study (Chan et al. (1991)) based on a sample of 479, 9-year old Chinese children in Hong Kong. The reaction time was measured using 12 attributes covering the components (1) movement time (MT): the time it takes to move the hand from one button to another, (2) simple reaction time (RT): the decision time in response to a signal, (3) choice reaction time: the decision time in response to more than one signal and (4) between-reaction times standard deviation over repeated tests on the same individual.

In the experiment, the fast reaction and movement times were assigned low scores, hence the negative correlation with intelligence means the positive association between reaction time and intelligence. Here the objective is to test the equality of the $k = 12$ correlation coefficients in the first row of the correlation matrix given in Table II, with $p = 13$.

The nuisance correlation coefficients are the $l = (13 \times 12)/2 - 12 = 66$, the correlation coefficients among the various measures of reaction and movement times. The median of the nuisance correlation coefficients is $r_* = 0.26$. The Q-statistic measuring the variability of Fisher transformed correlation coefficients in the first row of Table III is 6.2429 with $\bar{r}_W = -0.17$ based on $k = 12$ correlation coefficients. Substituting these numbers in the expression for C_1 and C_2 in Section 2.3 we get, $C_1 = 0.4927$ and $C_2 = 0.785$. Because $d = d_1 = 12 \times 11/2 = 66$, we have

$$\nu = 12 - 1 - 66 \times 0.4927/12 = 8.2902.$$

The resulting p-value is 0.6929. This p-value may be compared with 0.6868 obtained using an alternative procedure discussed in Meng et al. (1992), essentially the same.

3.2. EXAMPLE 3

For the second illustration we use the correlation coefficients given in Table II of Wheaton (1978) and is summarized in Table III. Measures of psychological (X_1) and psychophysiological (X_2) disorders are obtained at the baseline on $n = 603$ patients and these same measures are obtained sometime later (X_3 and X_4 respectively). Suppose that we wish to compare the correlation coefficient r_{14} (between X_1 and X_4) and r_{23} (between X_2 and X_3).

Table II. Correlation coefficients between several measures of reaction times and intelligence

Variable	1	2	3	4	5	6	7	8	9	10	11	12
Intelligence	−0.19	−0.19	−0.20	−0.14	−0.15	−0.17	−0.20	−0.18	−0.11	−0.10	−0.03	
1. Simple RT (mean)		0.68	0.33	0.49	0.23	0.18	0.50	0.45	0.31	0.27	0.17	0.03
2. Choice RT (mean)			0.53	0.38	0.41	0.31	0.41	0.68	0.22	0.18	0.18	−0.09
3. Odd-man-out (mean)				0.16	0.18	0.76	0.13	0.15	−0.01	0.17	0.49	0.06
4. Simple RT (SD)					0.58	0.30	0.27	0.24	0.10	0.46	0.50	0.08
5. Choice RT (SD)						0.35	0.14	0.12	−0.01	0.17	0.49	0.06
6. Odd-man-out RT (SD)							0.01	0.02	−0.12	0.15	0.19	−0.17
7. Simple MT (mean)								0.86	0.71	0.60	0.41	0.13
8. Choice MT (mean)									0.84	0.47	0.48	0.16
9. Odd-man-out MT (mean)										0.36	0.37	0.43
10. Simple MT (SD)											0.41	0.14
11. Choice MT (SD)												0.26
12. Odd-man-out MT (SD)												

Table III. Intercorrelations among four measures of disorder

	First occasion		Second occasion	
	Psychological disorder (X_1)	Psychophysical disorder (X_2)	Psychological disorder (X_3)	Psychophysical disorder (X_4)
First occasion: Psychological disorder (X_1)	–	0.45	0.53	0.38
First occasion: Psychophysical disorder (X_3)		–	0.25	0.31
Second occasion: Psychological disorder (X_3)			–	0.55
Second occasion: Psychophysical disorder (X_4)				–

Data abstracted from Table II of Wheaton (1978). The sociogenesis of psychological disorder: Reexamining the cusal issues with longitudinal data, *American Sociological Review* **43**, 383–403.

The null hypothesis of interest is H_o: $\rho_{14} = \rho_{23}$. The corresponding sample correlation coefficients are $r_{14} = 0.38$, $r_{23} = 0.25$, thus giving $\bar{z}_W = 0.325$ and $r_W = 0.314$. Also, from Table III, the median of the nuisance correlation coefficients is $r^* = 0.49$. The test statistic Q is 6.75. Here $d_1 = d = 1$, $k = 2$ and hence $C_1 = 0.45$ and $\nu = 0.55$ which yields $p = 0.0022$. This same data set is analyzed in Raghunathan et al. (1996) using an alternative method which yielded $p = 0.0014$.

4. Simulation Study

As described in the next section, the tests developed in this paper are based on asymptotic arguments and certain approximations. Consequently, we conducted a simulation experiment to find the actual levels of nominal 5% tests based on the two tests described above in specific situations. In the first simulation experiment, the objective was to test the null hypothesis of a common pairwise correlation coefficient among p variables based on a sample of size n. The data were generated from a multivariate normal distribution with mean zero, variance one and the common correlation coefficient ρ (i.e., the data were generated assuming the null hypothesis to be true). The conditions of the simulations were as follows:

- ρ = 0.1, 0.2, 0.3, 0.5 and 0.7
- p = 3, 4, 5, 7 and 10
- n = 25, 50, 75, 100, 150, 300 and 500

Thus, this simulation experiment can be considered as a $5 \times 5 \times 7$ factorial experiment. In each of the 175 cells, 10,000 data sets were generated and for each data set a 5% nominal test was constructed for the homogeneity of all the $p(p-1)/2$ correlation coefficients. The proportion of rejections as a percent of 10,000 replications was computed as an estimate of the actual level of the nominal 5% test. The levels of the test across all simulation conditions were between 3% and 7%. The actual levels were less than 5% for small p and large ρ and greater than 5% for small p and small ρ. The actual levels were then analyzed using the analysis of variance technique with 3 factors ρ, p and n. The factor n, the sample size, had negligible effect and the following regression model explained 94% of the variability in the actual levels,

$$\text{Actual Level} = 4.3167 - 5.8982\rho + 0.1510p + 0.5962\rho \times p.$$

The above equation may be useful in predicting the actual level of the nominal 5% test, given the common correlation coefficient estimate $\bar{r}_W$ and the dimensionality p.

In the second simulation experiment, the null hypothesis of interest was $H_o = \rho_{12} = \rho_{13} = \rho_{35} = \rho$ in a p-variate normal distribution. Here the simulation conditions were as follows:

- p = 5, 7 and 10
- ρ=0.1, 0.2, 0.3, 0.5 and 0.7 and
- n=25,50,75,100,150,300 and 500

Again 10,000 data sets were generated for each of 105 combinations of the factors from a multivariate normal distribution with mean zero, variance one and an arbitrary positive definite random correlation matrix except that $\rho_{12} = \rho_{21} = \rho_{13} = \rho_{31} = \rho_{35} = \rho_{53} = \rho$ (i.e., assuming that the null hypothesis is true and arbitrary values for the nuisance correlation coefficients). The nominal 5% tests were constructed for each data set as described in Section 2.3 and the actual level was defined as the percentage of rejections in 10,000 replications.

All the levels were between 4% and 6% and in the ANOVA of levels, none of the factors were significantly associated with the actual levels. Based on this simulation study, it seems that the simple chisquare tests for testing homogeneity of all or a subset of the product moment correlation coefficient has desired levels and should be adequate for most practical purposes.

5. Technical Justification

It is well known that, even for the modest size samples, Fisher r-to-z transform, $z = 0.5 \times \log[(1+r)/(1-r)]$ is approximately normally distributed with mean $z(\rho) = 0.5 \times \log[(1+\rho)/(1-\rho)]$ and variance $1/(n-3)$ where ρ is the population correlation coefficient and n is the sample size. Similarly, it can be shown, using the same arguments as in Olkin and Siotani (1976), that for a sample from a multivariate normal population and when the data are missing at random, $u^*_{ij} = \sqrt{(n_{ij} - 3)}(z_{ij} - z(\rho_{ij}))$ has a standard normal distribution with

$$C_1 = \mathrm{Cov}(u^*_{jk}, u^*_{jh}) = \frac{\rho_{kh}(1 - \rho^2_{jk} - \rho^2_{jh}) - \rho_{jk}\rho_{jh}(1 - \rho^2_{jk} - \rho^2_{jh} - \rho^2_{kh})/2}{(1 - \rho^2_{jk})(1 - \rho^2_{jh})}$$

and

$$\begin{aligned} C_2 \;=\; & \mathrm{Cov}(u^*_{jk}, u^*_{hm}) = \tfrac{1}{2(1-\rho^2_{jk})(1-\rho^2_{hm})}(\rho_{jh} - \rho_{jk}\rho_{kh})(\rho_{km} - \rho_{kh}\rho_{hm}) \\ & +(\rho_{jm} - \rho_{jh}\rho_{hm})(\rho_{kh} - \rho_{kj}\rho_{hj}) \\ & +(\rho_{jh} - \rho_{jm}\rho_{mh})(\rho_{km} - \rho_{kj}\rho_{hm}). \end{aligned}$$

Under the null hypothesis, $\rho_{ij} = \rho$ for all $i, j = 1, 2, \ldots, p; i \neq j$, the above two covariances simplify to

$$C_1 = \mathrm{Cov}(u^*_{jk}, u^*_{jh}) = \frac{\rho(3\rho + 2)}{2(1 + \rho)^2}$$

and

$$C_2 = \mathrm{Cov}(u^*_{jk}, u^*_{hm}) = 2\left(\frac{\rho}{1 + \rho}\right)^2.$$

The same asymptotic results hold when $z(\rho)$ in the definition of u^*_{ij} is replaced with its estimate, $\bar{z}_W = \sum_{ij}(n_{ij} - 3)z_{ij} / \sum_{ij}(n_{ij} - 3)$, (that is, for u_{ij} defined in Section 2). Further, it can shown that $Q = \sum_{ij} u^2_{ij}$ can be written as a quadaratic form U^tAU where U is a $d = p(p-1)/2$-dimensional vector $(u_{12}, u_{13}, \ldots, u_{p-1,p})^t$,

$$A = I_d - \frac{1}{d}E,$$

I_d is an identity matrix of order d, E is a $d \times d$ matrix of ones and the superscript t stands for matrix transpose. Let V denote a $d \times d$ covariance matrix of u_{ij} with diagonal elements equal to 1 and the off diagonal elements are either C_1 or C_2 depending upon whether the correlation is between overlapping or nonoverlapping correlation coefficients. The actual sampling distribution of the quadratic for Q is same as that of a linear combination of independent chisquare random variables. Instead, we approximate it by a chisquare random variable by matching the means of the two distributions. That is,

$$Q \approx \chi^2_\nu$$

where $\nu = E(Q) = tr(AV)$ where tr stands for the trace of the matrix (that is, the sum of its diagonal elements).

Now,

$$\mathrm{tr}(AV) = tr(V) - \mathrm{tr}(EV)/d = d - (d + 2d_1C_1 + 2d_2C_2)/d$$

where $d_1 = p(p-1)(p-2)/2$ is the number of overlapping pairs of correlation coefficients and $d_2 = p(p-1)(p-2)(p-3)/8$ is the number of nonoverlapping pairs of correlation coefficients. Algebraic simplification of the above expression reduces $tr(AV)$ to

$$d - 1 - \frac{\rho(p-2)(p\rho+2)}{(1+\rho)^2}.$$

Since ρ may not be specified, its estimate under the null $\bar{r}_W$ is substituted in the above expression to obtain ν defined in Section 2.1.

For deriving the test procedure for an arbitrary subset k of $p(p-1)/2$ possible pairwise correlation coefficients, we make a simplifying assumption that all the nuisance correlation coefficients are equal to ρ_*. Under this approximation,

$$C_1 = \frac{\rho^2\rho_*^2 + (2\rho_* - \rho^2)(1-2\rho^2)}{2(1-\rho^2)^2}$$

and

$$C_2 = \frac{2\rho_*^2}{(1+\rho)^2}.$$

Noting that V is $k \times k$ matrix and U is $k \times 1$ vector,

$$\mathrm{tr}(AV) = k - (k + 2d_1C_1 + 2d_2C_2)/k = k - 1 - 2\frac{d_1C_1 + d_2C_2}{k}$$

where d_1 and d_2 are the number of overlapping and nonoverlapping pairs of correlations coefficients respectively. Substituting the estimates $\bar{r}_W$ and r_* for ρ and ρ_* respectively, we obtain the results given in Section 2.3.

References

Brown, J. S. & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science* **2**(2), 155–192.

Bennet, B. M. (1978). On a test for equality of dependent correlation coefficients. *Statissche Hefte* **19**, 71–76.

Chan, J. W. C., Eysenck, H. J. & Lynn, R. (1991). Reaction times and intelligence among Hong Kong children. *Perceptual and Motor Skills* **72**, 427–433.

Choi, S. C. (1977). Tests for equality of dependent correlation coefficients. *Biometrika* **64**, 645–647.

Dunn, O. J. & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of American Statistical Association* **64**, 366–377.

Dunn, O. J. and Clark, V. A. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of American Statistical Association* **66**, 904–908.

Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Mathematical Statistics* **11**, 271–283.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Meng, X. L., Rosenthal, R. & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin* **111**, 172–175.

Neill, J. J. and Dunn, O. J. (1975). Equality of dependent correlation coefficients. *Biometrics* **31**, 531–543.

Olkin, I. and Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. In S. Ikeda (ed.), *Essays in Probability and Statistics*, Tokyo: Shinko Tersho.

Olkin, I. and Finn, J. D. (1990) Testing correlated correlations. *Psychological Bulletin* **108**, 330–333.

Olkin, I. & Finn, J. D. (1995). Correlation redux. *Psychological Bulletin* **118**, 155–164.

Pearson, K. & Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution. *Transactions of the Royal Society (London) Series A* **191**, 259–262.

Raghunathan, T. E., Rosenthal, R. & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods* **1**, 18–22.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin* **88**, 245–258.

Wheaton, B. (1978). The sociogenesis of psychological disorder: Reexamining the causal issues with longitudinal data. *American Sociological Review* **43**, 383–403.

Williams, E. J. (1959). The comparison of regression variables. *Journal of Royal Statistical Society, Series B* **21**, 396–399.