Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 2362–2379
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Inferences on correlation coefficients: One-sample, independent and correlated cases

K. Krishnamoorthy*, Yanping Xia

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA

Received 10 January 2006; received in revised form 26 July 2006; accepted 2 August 2006
Available online 10 November 2006

Abstract

This article concerns inference on the correlation coefficients of a multivariate normal distribution. Inferential procedures based on the concepts of generalized variables (GVs) and generalized p-values are proposed for elements of a correlation matrix. For the simple correlation coefficient, the merits of the generalized confidence limits and other approximate methods are evaluated using a numerical study. The study indicates that the proposed generalized confidence limits are uniformly most accurate even for samples as small as three. The results are extended for comparing two independent correlations, overlapping and non-overlapping dependent correlations. For each problem, the properties of the GV approach and other asymptotic methods are evaluated using Monte Carlo simulation. The GV approach produces satisfactory results for all the problems considered. The methods are illustrated using a few practical examples.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Coverage probabilities; Fisher's z transformation; Sizes; UMP test; Uniformly most accurate confidence interval

1. Introduction

The Pearson product–moment correlation is a standard measure of association between two continuous random variables. This correlation coefficient is also used in many other contexts. For example, the correlation coefficient between repeated measurements is used to judge the reliability of the measurement method. It is commonly used in the validation of psychological measures such as scales of anxiety and depression, where it is known as the test–retest reliability (Bland and Altman, 1996). Assuming normality, several authors have considered this problem and related problems, and provided solutions based on large-sample theory. We shall explain each of these problems with a brief review of literature in the following order.

1.1. Simple correlation

If the underlying variables are bivariate normally distributed, then an exact t procedure is available to test if the population correlation coefficient ρ is significantly different from zero. In situations where one wants to test for a non-zero value of ρ, Fisher's (1921) z transformation of the sample correlation coefficient R is a well-accepted method.

* Corresponding author.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2006.08.002

This approach provides reasonably accurate results for moderate samples, and software such as SAS and S-plus use this approach to find confidence limits for ρ. It is interesting to note that an exact method of computing confidence limits (similar to the Clopper–Pearson approach for computing binomial limits) is available in the literature (see Anderson, 1984, Section 4.2). It appears that this exact method is not popular even though it produces uniformly most accurate (UMA) confidence intervals. A reason could be that this exact method involves numerical integration of the probability density function of R (which is quite complex) and solving an integral equation, which pose some computational problems. Our experience suggests that some special computational techniques may be required to compute confidence limits using the exact method.

1.2. Comparing two independent correlations

The problem of comparing correlations of different groups arises in some practical situations. For example, it may be of interest to see if the correlation between the age and systolic blood pressure (SBP) for male adults differs from that for female adults. The following specific example is given in Bilker et al. (2004).

Example 1.1. This example concerns the correlation between a verbal memory score (v) and laterality of blood flow in each of three brain regions, namely, temporal (t), frontal (f) and subcortical (s). The data were measured on 14 men and 14 women, and the correlation matrix is given in Table 4. It is of interest, among others, to see if the correlations between the laterality of blood flow in each brain region and verbal memory score differ across gender.

Fisher's z transformation for the one-sample case can be readily extended to the problem of comparing two independent correlations. However, we note that this z transformation is useful for hypothesis testing, but the test cannot be transformed into a procedure for setting confidence limits. Olkin and Finn (OF) (Olkin and Finn, 1995) also proposed a normal-based asymptotic result that can be used for testing as well as obtaining confidence intervals. In general, the procedures given in the literature are based on asymptotic theory, and their validity for small samples is yet to be investigated.

1.3. Comparison of overlapping dependent correlations

This problem was originally motivated by the selection of the better of two available predictors (X_1 and X_2) for a dependent variable Y. In particular, one may want to compare the correlation between Y and X_1 with that of Y and X_2 to choose the better of these two predictors. This problem also arises in many other practical situations. For instance, in Example 1.1, it may be of interest to compare the correlation between the verbal memory score (Y) and the lateral blood flow in the temporal region (X_1) with that of Y and the lateral blood flow in the frontal region (X_2). Another example, given in OF (1990), is as follows.

Example 1.2. This example studies the correlations among measures related to cardiovascular health. The purpose of the study is to determine which of the cardiac measures (heart rate, SBP or diastolic blood pressure) is the best indicator of body mass index (BMI = weight/height^2), a general indicator of fitness among adult women. High values of BMI indicate severe obesity. If we are interested in comparing the corr(BMI, systolic pressure) with the corr(BMI, heart rate), then we have the problem of comparing two overlapping dependent correlations (as the BMI is a common variable). The correlation matrix for this example is given in Table 6.

Hotelling (1940) seems to be the first paper that addressed the problem of comparing two overlapping dependent correlations. He provided an exact conditional test; however, this test is not useful to infer the relationships in the population from which the sample is drawn (see Williams, 1959a). Williams (1959b) proposed an unconditional method which was obtained by modifying Hotelling's conditional test. Neill and Dunn (1975) compared 11 methods, including Williams' test, using Monte Carlo simulation. Based on the study, they concluded that the Williams test was the best. OF (1990) provided an asymptotic result which can be used for both hypothesis testing and obtaining confidence limits for the difference between two dependent correlations. Meng (MRR) (Meng et al., 1992) also provided an asymptotic method for hypothesis testing based on Fisher's z transformations of the sample correlation coefficients, but the test cannot be transformed into a procedure for setting confidence limits.

1.4. Comparison of two non-overlapping dependent correlations

The problem of comparing two non-overlapping dependent correlations has received some attention in the literature. Non-overlapping correlations arise when comparing the retest reliabilities of two measuring devices or comparing the correlation between two variables at two different time points (Raghunathan et al., 1996). OF (1990) provided the following specific example.

Example 1.3. Measures on SBP and BMI were collected from samples of mothers, youngest children and their older sisters. The correlation matrix is given in Table 8. The purpose of the study is to compare the correlations between SBP and BMI across the three groups. We here note that the correlations among these measures are interdependent but there is no common variable.

A test for comparing two non-overlapping dependent correlations dates back to Pearson and Filon (PF) Pearson and Filon (1898). This procedure has been extended to the problem of testing the equality of several correlations by OF (1990). PF method is based on the asymptotic distribution of the difference between the correlations to be compared. The test based on Fisher's z transformation is known to have better size properties than the PF test (see Raghunathan et al., 1996).

Our literature review showed that since Neill and Dunn's (1975) evaluation of the methods for comparing overlapping dependent correlations, no accuracy study or comparison study was made for the asymptotic procedures discussed above. Such studies are really warranted to identify the best of the available procedures. We also observe from the above review of literature that available procedures are applicable for large samples, and small-sample procedures (except for the simple correlation case) are yet to be developed. Furthermore, we note that the nature of the problems is very complex, and standard approaches may not yield any satisfactory results for small samples. Therefore, we investigate the applicability of the recent generalized variable (GV) approach (see Appendix A for a brief outline) for the aforementioned problems. The concept of generalized p-values for hypothesis testing has been introduced by Tsui and Weerahandi (1989) and of generalized confidence limits by Weerahandi (1993). For a detailed discussion and numerous applications, see the book by Weerahandi (1995a). This GV approach has been used successfully to develop tests and confidence intervals involving "non-standard" parameters, such as lognormal mean (Krishnamoorthy and Mathew, 2003) and normal quantiles (Weerahandi, 1995a). Applications of this approach include analysis of variance components (Zhou and Mathew, 1994), ANOVA under unequal variances (Weerahandi, 1995b), growth curve model (Weerahandi and Berger, 1999), common mean problem (Krishnamoorthy and Lu, 2003), tolerance limits for the one-way random effects model (Krishnamoorthy and Mathew, 2004) and multivariate Behrens–Fisher problem (Gamage et al., 2004).

In view of the above discussion, we develop inferential procedures based on the GV approach and compare them with the existing methods for each of the aforementioned problems. In the following section, we develop a generalized pivot variable for constructing confidence limits, and a generalized test variable for hypothesis testing of an element of the correlation matrix of a multivariate normal distribution. The GV approach is simple, and it can be easily applied to find solutions to all other problems discussed earlier. In Section 3, we outline the exact method, GV approach and other asymptotic approaches for making inference on a simple correlation. In Section 4, we consider the problem of comparing two independent correlations; GV method, and other asymptotic methods are provided. The GV approach and other methods for comparing two overlapping dependent correlations are given in Section 5, and the comparison of two non-overlapping dependent correlations is considered in Section 6. For each of the problems, the properties of the methods are evaluated using Monte Carlo simulation for small to large samples. Practical examples discussed earlier are used to illustrate the procedures. Some concluding remarks and applications of the GV approach to other problems are given in Section 7.

2. Generalized pivot variables for the elements of a covariance matrix

Let $X_1, \ldots, X_n$ be a sample from a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{S}$ denote the matrix of sums of squares and cross-products. That is,

$$\boldsymbol{S} = \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'. \tag{1}$$

It is well known that $\boldsymbol{S}=(s_{ij})$ follows a Wishart distribution with $n-1$ degrees of freedom and scale parameter matrix $\boldsymbol{\Sigma}=(\sigma_{ij})$. That is, $\boldsymbol{S}\sim W_p(n-1,\boldsymbol{\Sigma})$. Let $\boldsymbol{T}_s$ be the Cholesky factor of $\boldsymbol{S}$. That is, $\boldsymbol{S}=\boldsymbol{T}_s\boldsymbol{T}_s'$, where $\boldsymbol{T}_s=(T_{ij})$ is a lower triangular matrix with positive diagonal elements. Let $\boldsymbol{\theta}$ be the Cholesky factor of $\boldsymbol{\Sigma}$. Letting $\boldsymbol{V}=\boldsymbol{\theta}^{-1}\boldsymbol{T}_s=(V_{ij})$, we see that

$$\boldsymbol{V}\boldsymbol{V}'=\boldsymbol{\theta}^{-1}\boldsymbol{T}_s\boldsymbol{T}_s'\boldsymbol{\theta}'^{-1}\sim W_p(n-1,I_p), \tag{2}$$

where I_p is the identity matrix of order p. $\boldsymbol{V}$ is a lower triangular matrix and it is well known that the V_{ij}'s are all independent with

$$V_{ii}^2\sim\chi^2_{n-i},\quad i=1,\ldots,p\quad\text{and}\quad V_{ij}\sim \mathrm{N}(0,1),\quad i>j, \tag{3}$$

where χ^2_m denotes the chi-square random variable with m degrees of freedom (e.g., see Muirhead, 1982, p. 99).

Let $\boldsymbol{s}$ be an observed value of $\boldsymbol{S}$, and $\boldsymbol{t}_s$ be an observed value of $\boldsymbol{T}_s$ so that $\boldsymbol{t}_s\boldsymbol{t}_s'=\boldsymbol{s}$. Then

$$\boldsymbol{t}_s\boldsymbol{V}^{-1}=\boldsymbol{A}=(a_{ij})\text{ is a generalized pivot variable for }\boldsymbol{\theta}=(\theta_{ij}),\quad i\geqslant j. \tag{4}$$

Notice that $\boldsymbol{A}$ is a lower triangular matrix with $a_{ij}=0$ for $i<j$. The element a_{ij} is a generalized pivot variable for θ_{ij} for $i\geqslant j$. Furthermore, if $h(\boldsymbol{\theta})$ is a real-valued function of $\boldsymbol{\theta}$, then $h(\boldsymbol{A})$ is a generalized pivot variable for $h(\boldsymbol{\theta})$. For example, the correlation coefficient between the ith and jth components of a p-variate normal random vector can be expressed as

$$\rho_{ij}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}=\frac{\sum_{k=1}^{j}\theta_{ik}\theta_{jk}}{\sqrt{\sum_{k=1}^{i}\theta_{ik}^2}\sqrt{\sum_{k=1}^{j}\theta_{jk}^2}}=h(\boldsymbol{\theta})\quad\text{for } i\geqslant j. \tag{5}$$

The generalized pivot variable for $h(\boldsymbol{\theta})$ is given by

$$G_{\rho_{ij}}=\frac{\sum_{k=1}^{j}a_{ik}a_{jk}}{\sqrt{\sum_{k=1}^{i}a_{ik}^2}\sqrt{\sum_{k=1}^{j}a_{jk}^2}}=h(\boldsymbol{A})\quad\text{for } i\geqslant j. \tag{6}$$

Because, for a given $\boldsymbol{t}_s$, the distribution of $\boldsymbol{A}=\boldsymbol{t}_s\boldsymbol{V}^{-1}$ does not depend on any unknown parameters, the distribution of $h(\boldsymbol{A})$ can be obtained by Monte Carlo simulation. For example, the percentiles of $h(\boldsymbol{A})$ can be used to construct confidence intervals for $h(\boldsymbol{\theta})$. For more details and the requirements for a generalized pivot variable, see Appendix A.

The generalized pivot variable $G_{\rho_{ij}}$ can also be expressed in terms of the elements of the sample correlation matrix $\boldsymbol{r}=(r_{ij})$ based on $\boldsymbol{s}$. Notice that $\boldsymbol{r}=D_s^{-1}\boldsymbol{s}D_s^{-1}$, where $D_s=\mathrm{diag}(\sqrt{s_{11}},\ldots,\sqrt{s_{pp}})$. Let $\boldsymbol{t}_r$ be the Cholesky factor of $\boldsymbol{r}$, and define the lower triangular matrix $\boldsymbol{B}=\boldsymbol{t}_r\boldsymbol{V}^{-1}=(b_{ij})$. We can write the generalized pivot variable for ρ_{ij} as (see Appendix B)

$$G_{\rho_{ij}}=\frac{\sum_{k=1}^{j}b_{ik}b_{jk}}{\sqrt{\sum_{k=1}^{i}b_{ik}^2}\sqrt{\sum_{k=1}^{j}b_{jk}^2}}\quad\text{for } i>j. \tag{7}$$

For the purpose of hypothesis testing, the generalized test variable for ρ_{ij} is defined as

$$G^{\mathrm{t}}_{\rho_{ij}}=G_{\rho_{ij}}-\rho_{ij}\quad\text{for } i>j. \tag{8}$$

3. Inference on a simple correlation coefficient

We shall now consider the problem of testing and interval estimation of the population correlation coefficient ρ in a bivariate normal setup. Let R denote the sample correlation coefficient based on $\boldsymbol{S}$ in (1). That is,

$$R=\frac{S_{21}}{\sqrt{S_{11}S_{22}}}\quad\text{and}\quad\rho=\frac{\sigma_{21}}{\sqrt{\sigma_{11}\sigma_{22}}}, \tag{9}$$

where S_{ij} is the (i,j)th element of $\boldsymbol{S}$ in (1). Let r be an observed value of R.

The following exact inferential procedures for ρ are based on n and r.

3.1. Exact methods

Suppose we want to test

$$\mathrm{H}_0 : \rho \leqslant \rho_0 \text{ vs. } \mathrm{H}_a : \rho > \rho_0, \tag{10}$$

where ρ_0 is a specified value of ρ. Then, for an observed r and the nominal level α, the test that rejects H_0 whenever the p-value, $P(R > r|n, \rho_0)$, is less than α is a uniformly most powerful (UMP) invariant test (Anderson, 1984, p. 114). The p-value for $\mathrm{H}_0 : \rho = \rho_0$ vs. $\mathrm{H}_a : \rho \neq \rho_0$ is given by $2\min\{P(R > r|n, \rho_0), P(R \leqslant r|n, \rho_0)\}$.

The acceptance regions (over the parameter space) form exact confidence intervals. The endpoints of the $100(1-\alpha)\%$ confidence interval can be obtained as the solutions of some integral equations. As detailed in Anderson (1984, Section 4.2), the upper limit ρ_{U} for ρ is the solution of the equation

$$P(R \leqslant r|n, \rho_{\mathrm{U}}) = \alpha/2, \tag{11}$$

and the lower limit ρ_{L} is the solution of the equation

$$P(R \geqslant r|n, \rho_{\mathrm{L}}) = \alpha/2. \tag{12}$$

One-sided $100(1-\alpha)\%$ lower limits can be obtained by replacing $\alpha/2$ by α in Eqs. (11) and (12). These confidence limits are UMA, because they are the acceptance regions of the UMP tests (see Casella and Berger, 2002, p. 445).

The p-values of the exact tests can be computed accurately (as long as ρ_0 is not close to its boundaries and n is not large) by numerically integrating the probability density function of R given in Hotelling (1953). However, as already mentioned, solving Eqs. (11) and (12) seems to be difficult.

3.2. Approximate methods for a simple correlation

3.2.1. Fisher's z transformation

Fisher's tests and confidence intervals are based on the following asymptotic distribution. Let $Z = \tanh^{-1}(R) = \frac{1}{2}\ln((1+R)/(1-R))$ and $\mu_\rho = \tanh^{-1}(\rho) = \frac{1}{2}\ln((1+\rho)/(1-\rho))$. Then

$$Z \sim \mathrm{N}(\mu_\rho, (n-3)^{-1}) \text{ asymptotically.} \tag{13}$$

As μ_ρ is an increasing function of ρ, inferential procedures about ρ can be obtained from the above asymptotic distribution. Specifically, let z be an observed value of Z. That is, $z = \frac{1}{2}\ln((1+r)/(1-r))$. Then, an approximate $100(1-\alpha)\%$ confidence interval for ρ is given by $(\tanh[z - z_{\alpha/2}/\sqrt{n-3}], \tanh[z + z_{\alpha/2}/\sqrt{n-3}])$, where z_p is the upper pth quantile of the standard normal distribution. Furthermore, the p-value for testing (10) is given by $P(Z > z|n, \rho_0) = 1 - \Phi(\sqrt{n-3}(z - \mu_{\rho_0}))$, where Φ is the standard normal distribution function.

3.2.2. Jayaratnam's (1992) approach

Jayaratnam proposed an approximate confidence interval for ρ that is given by

$$\left(\frac{r-w}{1-rw}, \frac{r+w}{1+rw}\right), \tag{14}$$

where $w = \left[t_{n-2,1-\alpha/2}/\sqrt{n-2}\right] / \sqrt{1 + [t_{n-2,1-\alpha/2}]^2/(n-2)}$ and $t_{m,p}$ denotes the pth quantile of Student's t distribution with $\mathrm{d}f = m$.

3.3. Generalized test and confidence limit for ρ

Notice that for the bivariate case, the generalized pivot variable for ρ is $G_{\rho_{21}}$ in (7). After simplification, we get

$$G_\rho = G_{\rho_{21}} = \frac{b_{21}}{\sqrt{b_{21}^2 + b_{22}^2}}, \tag{15}$$

where the b_{ij}'s are the elements of

$$\boldsymbol{B} = \boldsymbol{t}_r \boldsymbol{V}^{-1} = \begin{pmatrix} 1 & 0 \\ r & \sqrt{1-r^2} \end{pmatrix} \begin{pmatrix} \dfrac{1}{V_{11}} & 0 \\ -\dfrac{V_{21}}{V_{11}V_{22}} & \dfrac{1}{V_{22}} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{V_{11}} & 0 \\ \dfrac{r}{V_{11}} - \dfrac{V_{21}\sqrt{1-r^2}}{V_{11}V_{22}} & \dfrac{\sqrt{1-r^2}}{V_{22}} \end{pmatrix}.$$

Using the elements of $\boldsymbol{B}$ in (15), and after some simplification, we get

$$G_\rho = \frac{r^* V_{22} - V_{21}}{\sqrt{(r^* V_{22} - V_{21})^2 + V_{11}^2}} \quad \text{where } r^* = r/\sqrt{1-r^2}. \tag{16}$$

As we mentioned earlier, for a given r^*, the distribution of G_ρ does not depend on any unknown parameters, and hence its distribution or percentiles can be estimated using Monte Carlo simulation. In particular, the αth and $(1-\alpha)$th quantiles of G_ρ form a $100(1-2\alpha)\%$ confidence interval for ρ.

The generalized p-value for testing hypotheses in (10) is given by $P(G_\rho \leqslant \rho_0)$ (see Appendix A), and it can be estimated by Monte Carlo simulation. The generalized test rejects the H_0 in (10) whenever the generalized p-value is less than the nominal level.

The following algorithm, which is based on (16) and the distributional results of V_{ij} in (3), can be used to estimate the generalized confidence limits and the generalized p-values.

Algorithm 1.

For a given n and r:
  Set $r^* = r/\sqrt{1-r^2}$
For $i = 1$ to m
  Generate $Z_0 \sim N(0, 1)$, $U_1 \sim \chi^2_{n-1}$ and $U_2 \sim \chi^2_{n-2}$
  Set $Q_i = \dfrac{r^*\sqrt{U_2} - Z_0}{\sqrt{(r^*\sqrt{U_2} - Z_0)^2 + U_1}}$
[end loop]

The 100αth percentile of the Q_i's is a $100(1-\alpha)\%$ lower limit for ρ. The generalized p-value for testing hypotheses in (10) is estimated by the proportion of Q_i's that are less than or equal to ρ_0.

3.4. Accuracy studies of inferential procedures for ρ

To appraise the small-sample properties of the generalized inferential methods, we computed 95% upper limits for ρ when $n = 3$ and r ranges from 0 to 0.95. These upper limits ρ_U are computed using Algorithm 1 with $m = 1,000,000$ and are given in Table 1 along with the probabilities computed using (11). For a given r, accuracy of the limit ρ_U can be judged by the closeness of $P(R \leqslant r|n, \rho_U)$ to 0.05. Recall that for the UMA limit, this probability must be 0.05. We see from Table 1 that these probabilities are equal to 0.05 for all cases considered. Thus, we see that the generalized limits practically coincide with the UMA limits even for samples of size 3.

In Table 1, we present 95% upper limits based on the GV approach, Jayaratnam's (1992) method (J) and Fisher's z transformation (F). We also present the probability in (11) for each limit (the number in parentheses). We again observe from the table values that these probabilities for the generalized limits are equal to 0.05 for all the cases considered. For $n \geqslant 5$, the limits based on Jayaratnam's approach and Fisher's z transformation are almost the same for all values of r considered. For these limits, the probabilities in (11) are smaller than 0.05 indicating that Jayaratnam's method and Fisher's method are conservative for small to moderate samples.

3.5. An illustrative example for a simple correlation coefficient

We shall now illustrate the results for the simple correlation coefficient using the following example.

Table 1
(a) The 95% generalized upper limits for ρ when $n = 3$, $P = P(R \leqslant r|n, \rho_U)$ and (b) 95% upper limits ρ_U for ρ

(a)											
r	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
ρ_U	0.906	0.911	0.922	0.932	0.942	0.952	0.961	0.970	0.980	0.990	0.995
P	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)

(b)												
r	$n = 5$			$n = 10$			$n = 20$			$n = 30$		
	GV	J	F	GV	J	F	GV	J	F	GV	J	F
0.00	0.729	0.805	0.822	0.521	0.549	0.552	0.368	0.378	0.379	0.301	0.306	0.306
	(0.05)[a]	(0.03)	(0.02)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
0.05	0.749	0.822	0.838	0.556	0.583	0.586	0.410	0.420	0.421	0.344	0.351	0.351
	(0.05)	(0.03)	(0.02)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
0.30	0.833	0.890	0.900	0.702	0.729	0.731	0.597	0.609	0.610	0.547	0.555	0.555
	(0.05)	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
0.60	0.912	0.948	0.952	0.845	0.864	0.865	0.788	0.797	0.798	0.759	0.765	0.766
	(0.05)	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
0.80	0.958	0.976	0.979	0.927	0.937	0.938	0.899	0.905	0.905	0.885	0.889	0.889
	(0.05)	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
0.95	0.990	0.994	0.995	0.982	0.985	0.985	0.976	0.977	0.977	0.972	0.973	0.973
	(0.05)	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)

[a]The numbers in parentheses are $P(R \leqslant r|n, \rho_U)$.

Example 3.1. The data for this example are taken from Fisher and Van Belle (1993, Example 9.3). The data represent *erythrocyte adenosine triphosphate* (ATP) levels in youngest and oldest sons in 17 families. For easy reference, the data are given in Table 2. The ATP level is important because it determines the ability of the blood to carry energy to the cells of the body.

The 95% confidence intervals based on the first five pairs, the first 10 pairs and the whole data set are given in Table 2. We observe from the table that the results are in agreement with the numerical properties that we studied earlier. Specifically, for smaller sample sizes, Fisher's z and Jayaratnam's methods are conservative, and so they produced wider intervals than the generalized confidence intervals. The differences among the confidence intervals based on the entire data are not appreciable.

4. Comparison between two independent correlations

Let $X_{k1}, \ldots, X_{kn_k}$ be a bivariate normal sample from $N_2(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $k = 1, 2$. Define

$$\boldsymbol{S}_k = \sum_{j=1}^{n_k} (\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)(\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)', \quad k = 1, 2. \tag{17}$$

The sample and population correlation coefficients are defined, respectively, by

$$R_k = \frac{S_{k,12}}{\sqrt{S_{k,11}S_{k,22}}} \quad \text{and} \quad \rho_k = \frac{\sigma_{k,12}}{\sqrt{\sigma_{k,11}\sigma_{k,22}}},$$

where $S_{k,ij}$ is the (i, j)th element of $\boldsymbol{S}_k$ and $\sigma_{k,ij}$ is the (i, j)th element of $\boldsymbol{\Sigma}_k$, for $k = 1, 2$. Furthermore, let r_k be an observed value of R_k, $k = 1, 2$.

4.1. Fisher's z transformation for comparing two independent correlations

The one-sample Fisher's z transformation can be easily extended to the two-sample case. Define $Z_k = \frac{1}{2}\ln((1 + R_k)/(1 - R_k))$ and $\mu_{\rho_k} = \frac{1}{2}\ln((1 + \rho_k)/(1 - \rho_k))$, $k = 1, 2$. Since R_1 and R_2 are independent, it follows from the

Table 2
(a) ATP levels in youngest and oldest sons and (b) 95% confidence intervals for correlations between ATP levels in youngest and oldest sons and other relevant statistics

Family	Youngest (x)	Oldest (y)
(a)		
1	4.18	4.81
2	5.16	4.18
3	4.85	4.48
4	3.43	4.19
5	4.53	4.27
6	5.13	4.87
7	4.10	4.74
8	4.77	4.53
9	4.12	3.72
10	4.65	4.62
11	6.03	5.83
12	5.94	4.40
13	5.99	4.87
14	5.43	5.44
15	5.00	4.70
16	4.82	4.14
17	5.25	5.30

n	r	Generalized limits[a]	Fisher's z	Jayaratnam (1992)
(b)				
5	0.606	$r^* = 0.761$; (−0.501, 0.942)	$z = 0.702$; (−0.594, 0.970)	$w = 0.879$; (−0.583, 0.969)
10	0.544	$r^* = 0.648$; (−0.126, 0.856)	$z = 0.610$; (−0.130, 0.874)	$w = 0.632$; (−0.134, 0.875)
17	0.597	$r^* = 0.745$; (0.156, 0.827)	$z = 0.689$; (0.164, 0.838)	$w = 0.482$; (0.162, 0.838)

[a]Computed using Algorithm 1 with 1,000,000 runs.

asymptotic result in (13) that

$$\frac{(Z_1 - Z_2) - (\mu_{\rho_1} - \mu_{\rho_2})}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}} \sim \mathrm{N}(0, 1) \text{ asymptotically.} \tag{18}$$

Using the above asymptotic result one can easily develop test procedures for $\rho_1 - \rho_2$. Specifically, for an observed value (z_1, z_2) of (Z_1, Z_2), the p-value for testing $\mathrm{H}_0 : \rho_1 \leqslant \rho_2$ vs. $\mathrm{H}_a : \rho_1 > \rho_2$ is given by $1 - \Phi((z_1 - z_2)/\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)})$, where Φ is the standard normal distribution function. Notice that, using the distributional result in (18), one can easily obtain a confidence interval for $\mu_{\rho_1} - \mu_{\rho_2}$ but not for $\rho_1 - \rho_2$.

4.2. *OF (1995) method for comparing two independent correlations*

OF (1995) proposed an asymptotic distribution for $R_1 - R_2$ which can be used for hypothesis testing and setting confidence limits for $\rho_1 - \rho_2$. Specifically, their test is based on the result that

$$\frac{(R_1 - R_2) - (\rho_1 - \rho_2)}{\sqrt{\mathrm{var}(R_1) + \mathrm{var}(R_2)}} \sim \mathrm{N}(0, 1) \text{ asymptotically,} \tag{19}$$

where the large-sample variance $\mathrm{var}(R_i) = (1 - R_i^2)^2/n_i$, $i = 1, 2$. For a given observed value (r_1, r_2) of (R_1, R_2), the $100(1 - \alpha)\%$ confidence interval for $\rho_1 - \rho_2$ is given by

$$r_1 - r_2 \pm z_{\alpha/2}\sqrt{\frac{(1 - r_1^2)^2}{n_1} + \frac{(1 - r_2^2)^2}{n_2}},$$

where z_p denotes the upper pth quantile of the standard normal distribution.

4.3. GV approach for comparing two independent correlations

The generalized pivot variable for the two-sample case can be obtained from the one for the one-sample case in a straightforward manner. If G_{ρ_k} is a generalized pivot variable for ρ_k, $k = 1, 2$, then the generalized pivot variable for $\rho_1 - \rho_2$ is $G_{\rho_1} - G_{\rho_2}$. Letting $r_k^* = r_k/\sqrt{1 - r_k^2}$, $k = 1, 2$, we can write the generalized pivot variable for $\rho_1 - \rho_2$ as

$$\begin{aligned} G_{\rho_1-\rho_2} &= G_{\rho_1} - G_{\rho_2} \\ &= \frac{r_1^* V_{1,22} - V_{1,21}}{\sqrt{\left(r_1^* V_{1,22} - V_{1,21}\right)^2 + V_{1,11}^2}} - \frac{r_2^* V_{2,22} - V_{2,21}}{\sqrt{\left(r_2^* V_{2,22} - V_{2,21}\right)^2 + V_{2,11}^2}}, \end{aligned} \tag{20}$$

where the $V_{k,ij}$'s are independent random variables with

$$V_{k,11}^2 \sim \chi_{n_k-1}^2, \quad V_{k,22}^2 \sim \chi_{n_k-2}^2 \quad \text{and} \quad V_{k,21} \sim \mathrm{N}(0, 1), \quad k = 1, 2.$$

For a given r_1^* and r_2^*, the distribution of the generalized pivot variable in (20) does not depend on any unknown parameters, and so its percentiles can be estimated using Monte Carlo simulation as in the one-sample case. Appropriate percentiles form confidence limits for $\rho_1 - \rho_2$. In particular, lower and upper 2.5th percentiles of $G_{\rho_1} - G_{\rho_2}$ are the endpoints of the 95% generalized confidence interval.

The generalized test variable for $\rho_1 - \rho_2$ is $G^{\mathrm{t}}_{\rho_1-\rho_2} = G_{\rho_1} - G_{\rho_2} - (\rho_1 - \rho_2)$. For fixed r_1^* and r_2^*, this generalized test variable is stochastically decreasing in $\rho_1 - \rho_2$, and so the generalized p-value for testing $\mathrm{H}_0 : \rho_1 - \rho_2 \leqslant c$ vs. $\mathrm{H}_a : \rho_1 - \rho_2 > c$, where c is a specified number, is given by

$$\sup_{\mathrm{H}_0} P(G^{\mathrm{t}}_{\rho_1-\rho_2} \leqslant 0) = \sup_{\mathrm{H}_0} P(G_{\rho_1} - G_{\rho_2} \leqslant \rho_1 - \rho_2) = P(G_{\rho_1} - G_{\rho_2} \leqslant c).$$

The above generalized p-value can be estimated using Monte Carlo simulation as in the one-sample case.

4.4. Monte Carlo studies for independent correlations

The sizes of Fisher's test, OF (1995) test and the generalized test for the independent case are estimated using Monte Carlo simulation. The simulation study for the generalized test is carried out as follows: for a given (n_1, n_2) and (ρ_1, ρ_2), we first generated 2500 pairs of (r_1, r_2). For each generated (r_1, r_2), we used a simulation consisting of 5000 runs to estimate the generalized p-value. The proportion of the 2500 generalized p-values which are less than the nominal level α is an estimate of the Type I error rate. The necessary chi-square variates and normal variates are generated using IMSL subroutines RNCHI and RNNOA, respectively. For Fisher's test and OF test, we used simulation consisting of 100,000 runs to estimate the sizes.

The Type I error rates provided in Table 3 are estimated for equal sample sizes 5, 10, 15, 20, 30 and 50. We observe from Table 3 that the sizes of the generalized test are slightly larger (liberal) than the nominal level for small samples, and they are very close to the nominal level for moderate samples. On the other hand, Fisher's test has smaller sizes (conservative) when samples are small; otherwise its sizes are close to the nominal level. OF test appears to be too liberal when the common value of ρ_1 and ρ_2 is small, and too conservative when it is large. This behavior seems to diminish with increasing sample sizes (see the case $n_1 = n_2 = 50$).

The coverage properties of the generalized confidence limits for $\rho_1 - \rho_2$ can be inferred from its size properties. In particular, the coverage probabilities are expected to be slightly smaller than the nominal level for very small samples, and close to the nominal level for moderate to large samples.

4.5. An example for comparing two independent correlations

Example 1.1 (*Continued*). Based on samples of 14 men and 14 women, the correlations between a verbal memory score (v) and laterality of blood flow in each of three brain regions, namely, temporal (t), frontal (f) and subcortical (s) are given in Table 4. We estimated the 90% lower limits and p-values for hypothesis tests using the three methods of this section and provided them in Table 4. We see from Table 4 that Fisher's test and the GV test produced practically the same p-values whereas the OF approach gives a slightly smaller p-value. As Fisher's method is not applicable for

Table 3
Sizes of the tests for comparing two independent correlations $H_0 : \rho_1 \leqslant \rho_2$ vs. $H_0 : \rho_1 > \rho_2$; $\alpha = 0.05$

$\rho_1 = \rho_2$	$(n_1, n_2) = (5, 5)$			$(n_1, n_2) = (10, 10)$			$(n_1, n_2) = (15, 15)$		
	GV	F	OF	GV	F	OF	GV	F	OF
0.00	0.05	0.05	0.13	0.05	0.05	0.09	0.05	0.05	0.08
0.05	0.05	0.04	0.14	0.05	0.05	0.09	0.04	0.05	0.08
0.10	0.05	0.04	0.14	0.06	0.05	0.09	0.05	0.05	0.08
0.30	0.05	0.04	0.12	0.06	0.05	0.09	0.05	0.05	0.07
0.50	0.05	0.04	0.10	0.05	0.05	0.07	0.05	0.05	0.06
0.80	0.05	0.04	0.04	0.05	0.05	0.03	0.05	0.05	0.04
0.95	0.06	0.04	0.01	0.05	0.05	0.00	0.05	0.05	0.02
$\rho_1 = \rho_2$	$(n_1, n_2) = (20, 20)$			$(n_1, n_2) = (30, 30)$			$(n_1, n_2) = (50, 50)$		
	GV	F	OF	GV	F	OF	GV	F	OF
0.00	0.06	0.05	0.07	0.05	0.05	0.06	0.06	0.05	0.06
0.05	0.05	0.05	0.07	0.05	0.05	0.06	0.06	0.05	0.06
0.10	0.05	0.05	0.07	0.05	0.05	0.06	0.05	0.05	0.06
0.30	0.05	0.05	0.07	0.05	0.05	0.06	0.05	0.05	0.06
0.50	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.05
0.80	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.05
0.95	0.05	0.05	0.02	0.05	0.05	0.03	0.05	0.05	0.04

GV: Generalized variable; F: Fisher's z transformation; OF: Olkin–Finn test.

Table 4
(a) Correlations between laterality of blood flow in three brain regions and verbal memory score and (b) p-values of the tests and 90% lower limits for comparing independent correlations in (a)

Gender	n	Blood flow rate laterality		
		Temporal	Subcortical	Frontal
(a)				
Male	14	$r_{M,vt} = -0.340$	$r_{M,vs} = +0.641$	$r_{M,vf} = -0.032$
Female	14	$r_{F,vt} = +0.812$	$r_{F,vs} = +0.491$	$r_{F,vf} = -0.212$

Method	p-values			Lower limits		
	(1)	(2)	(3)	$\rho_{F,vt} - \rho_{M,vt}$	$\rho_{M,vs} - \rho_{F,vs}$	$\rho_{M,vf} - \rho_{F,vf}$
(b)						
GV	0.000	0.298	0.331	0.716	−0.207	−0.324
Fisher	0.000	0.301	0.334	–	–	–
Olkin–Finn	0.000	0.280	0.313	0.827	−0.179	−0.293

(1) $H_a : \rho_{F,vt} > \rho_{M,vt}$; (2) $H_a : \rho_{M,vs} > \rho_{F,vs}$; (3) $H_a : \rho_{M,vf} > \rho_{F,vf}$.

finding confidence limits, we presented 90% lower limits based on the GV method and OF method. These two methods produced lower limits that are appreciably different.

5. Comparison of two overlapping dependent correlations

We shall now consider the problem of comparing two overlapping dependent correlations ρ_{21} and ρ_{31} in the correlation matrix

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{21} & \rho_{31} \\ \rho_{21} & 1 & \rho_{32} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix},$$

based on a sample correlation matrix $\boldsymbol{R} = (R_{ij})$.

5.1. William's (1959b) method for comparing two overlapping dependent correlations

William's modification to the Hotelling's test (as given in Neill and Dunn, 1975) is based on the asymptotic result that

$$Z_{\mathrm{WH}} = (R_{21} - R_{31})\left(\frac{(n-1)(1+R_{32})}{2((n-1)/(n-3))|\boldsymbol{R}| + \bar{R}^2(1-R_{32})^3}\right)^{1/2} \sim t_{n-3} \quad \text{asymptotically,} \tag{21}$$

where $\bar{R} = (R_{21} + R_{31})/2$ and $|\boldsymbol{R}| = 1 - R_{21}^2 - R_{31}^2 - R_{32}^2 + 2R_{21}R_{31}R_{32}$ is the determinant of $\boldsymbol{R}$. Inferential procedures for $\rho_{21} - \rho_{31}$ can be readily obtained from the distributional result in (21).

5.2. OF (1990) method for comparing two overlapping dependent correlations

OF approach uses the result that

$$\frac{(R_{21} - R_{31}) - (\rho_{21} - \rho_{31})}{\sqrt{\mathrm{var}(R_{21} - R_{31})}} \sim \mathrm{N}(0, 1) \text{ asymptotically,}$$

where $\mathrm{var}(R_{21} - R_{31}) = \mathrm{var}(R_{21}) + \mathrm{var}(R_{31}) - 2\,\mathrm{cov}(R_{21}, R_{31})$, $\mathrm{var}(R_{i1}) = (1 - R_{i1}^2)^2/n$, $i = 2, 3$ and

$$\mathrm{cov}(R_{21}, R_{31}) = [0.5(2R_{32} - R_{21}R_{31})(1 - R_{21}^2 - R_{31}^2 - R_{32}^2) + R_{32}^3]/n.$$

5.3. MRR (1992) test for comparing two overlapping dependent correlations

MRR (1992) proposed a test based on the following asymptotic result. Let Z_{21} and Z_{31} denote the Fisher's z transformations (see Section 3) of R_{21} and R_{31}, respectively. Then, the statistic

$$\frac{Z_{21} - Z_{31}}{\sqrt{2(1-R_{32})q/(n-3)}} \sim \mathrm{N}(0, 1) \quad \text{asymptotically,} \tag{22}$$

where

$$q = \frac{1 - f(R_{21}^2 + R_{31}^2)/2}{1 - (R_{21}^2 + R_{31}^2)/2} \quad \text{and} \quad f = \min\left\{\frac{1 - R_{32}}{2(1 - (R_{21}^2 + R_{31}^2)/2)}, 1\right\}.$$

5.4. The GV approach for comparing two overlapping dependent correlations

The GV for $\rho_{21} - \rho_{31}$ can be obtained from (7) in a straightforward manner. In particular,

$$\begin{aligned} G_{\rho_{21}-\rho_{31}} &= G_{\rho_{21}} - G_{\rho_{31}} \\ &= \frac{b_{21}}{\sqrt{b_{21}^2 + b_{22}^2}} - \frac{b_{31}}{\sqrt{b_{31}^2 + b_{32}^2 + b_{33}^2}}. \end{aligned} \tag{23}$$

Recall that the b_{ij}'s are the elements of $\boldsymbol{B} = \boldsymbol{t}_r\boldsymbol{V}^{-1}$ and that $\boldsymbol{t}_r\boldsymbol{t}_r' = \boldsymbol{r}$. After some calculation, we can express

$$G_{\rho_{21}} = \frac{r_{21}^* V_{22} - V_{21}}{\sqrt{(r_{21}^* V_{22} - V_{21})^2 + V_{11}^2}} \quad \text{where } r_{21}^* = r_{21} \Big/ \sqrt{1 - r_{21}^2}\,.$$

To find $G_{\rho_{31}}$, we can explicitly express the elements of $\boldsymbol{B}$ as

$$b_{31} = \frac{r_{31}}{V_{11}} - \frac{V_{21}(r_{32} - r_{21}r_{31})}{V_{11}V_{22}\sqrt{1-r_{21}^2}} + \frac{(V_{21}V_{32} - V_{22}V_{31})}{V_{11}V_{22}V_{33}}\sqrt{\frac{|\boldsymbol{r}|}{1-r_{21}^2}},$$

$$b_{32} = \frac{r_{32} - r_{21}r_{31}}{V_{22}\sqrt{1-r_{21}^2}} - \frac{V_{32}}{V_{22}V_{33}}\sqrt{\frac{|\boldsymbol{r}|}{1-r_{21}^2}} \quad \text{and} \quad b_{33} = \frac{1}{V_{33}}\sqrt{\frac{|\boldsymbol{r}|}{1-r_{21}^2}}.$$

The distributions of the V_{ij}'s are given in (3).

For a given r_{21}, r_{31} and r_{32}, the distribution of the generalized pivot variable $G_{\rho_{21}} - G_{\rho_{31}}$ does not depend on any unknown parameters, and so its percentiles can be estimated using Monte Carlo simulation. Appropriate percentiles form confidence limits for $\rho_{21} - \rho_{31}$. The generalized p-value for testing

$$\mathrm{H}_0 : \rho_{21} \leqslant \rho_{31} \text{ vs. } \mathrm{H}_a : \rho_{21} > \rho_{31}$$

is given by $P(G_{\rho_{21}} \leqslant G_{\rho_{31}})$, which can be estimated by simulation.

Remark 5.1. In some situations, one may want to compare the squared correlation coefficients. For example, if Y and X_1 are negatively correlated and Y and X_2 are positively correlated, then ρ_{21}^2 and ρ_{31}^2 should be compared to determine the better predictor. In this case, the generalized test can be readily obtained because the generalized pivot variable for ρ_{ij}^2 is $(G_{\rho_{ij}})^2$. Specifically, the generalized p-value for testing $\mathrm{H}_0 : \rho_{21}^2 \leqslant \rho_{31}^2$ vs. $\mathrm{H}_a : \rho_{21}^2 > \rho_{31}^2$ is $P((G_{\rho_{21}})^2 \leqslant (G_{\rho_{31}})^2)$, which can be estimated using simulation. If the sample correlations have opposite signs, then the asymptotic tests can be applied after changing the sign of the negative correlation and the sign of r_{32}.

5.5. Monte Carlo studies for comparing two overlapping dependent correlations

We carried out simulation studies along the lines of Section 4.4 to understand the size properties of the generalized test and other asymptotic methods. For simulation studies, we can take $\boldsymbol{\Sigma}$ as a correlation matrix without loss of generality. For $p = 3$, we estimated the sizes of the GV test, Williams (1959b) test (W), OF test and MRR test for

$$\mathrm{H}_0 : \rho_{21} \leqslant \rho_{31} \text{ vs. } \mathrm{H}_a : \rho_{21} > \rho_{31},$$

when $n = 10$, 20, 30 and 40. The estimated sizes are given in Table 5. We observe the following from the table values. (i) The MRR and Williams' tests control Type I error rates satisfactorily in all situations considered. (ii) The GV method performs well only when the sample sizes are 30 or more. (iii) The OF test is, in general, liberal for small and moderate samples.

As Williams' method can be used for both hypothesis testing and setting confidence limits, and it performs as well as the MRR test (which cannot be easily transformed to obtain confidence limits), we recommend Williams' method for applications.

5.6. An example for comparing two overlapping dependent correlations

Example 1.2 (*Continued*). The correlation matrix of cardiovascular measures and BMI of a sample of 66 black mothers (who have children 11.7 years old or younger) is given in Table 6.

The p-values and 90% lower limits for comparing different pairs of correlations are given in Table 6. Notice that the MRR test is not easy to invert to obtain a confidence interval. As the sample size is considerably large, all the tests produced practically the same p-values. The GV approach, William's method and the OF method all produced confidence limits which are not appreciably different. Even though the pairs of correlations are compared one at a time, SBP seems to be the best indicator of the BMI for adult black women.

Table 5
Sizes of the tests for testing two overlapping dependent correlations $H_0 : \rho_{21} \leqslant \rho_{31}$ vs. $H_a : \rho_{21} > \rho_{31}$; $\alpha = 0.05$

ρ_{32}	$\rho_{21} = \rho_{31}$	$n = 10$				$n = 20$			
		GV	W	OF	MRR	GV	W	OF	MRR
−0.60	0.10	0.06	0.04	0.11	0.05	0.05	0.05	0.08	0.05
	0.20	0.05	0.05	0.10	0.05	0.05	0.05	0.08	0.05
	0.40	0.04	0.04	0.09	0.05	0.04	0.05	0.07	0.05
0.00	0.10	0.05	0.05	0.10	0.05	0.05	0.05	0.07	0.05
	0.40	0.06	0.04	0.08	0.05	0.06	0.05	0.07	0.05
	0.60	0.05	0.04	0.06	0.05	0.05	0.05	0.06	0.05
	0.70	0.05	0.04	0.04	0.05	0.06	0.05	0.05	0.05
0.30	0.10	0.05	0.05	0.09	0.04	0.06	0.05	0.07	0.04
	0.40	0.06	0.05	0.08	0.04	0.07	0.05	0.07	0.05
	0.60	0.07	0.05	0.06	0.05	0.07	0.05	0.06	0.05
		$n = 30$				$n = 40$			
−0.60	0.10	0.05	0.05	0.07	0.05	0.05	0.05	0.06	0.05
	0.20	0.05	0.05	0.07	0.05	0.06	0.05	0.06	0.05
	0.40	0.06	0.05	0.06	0.05	0.05	0.05	0.06	0.05
0.00	0.10	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.05
	0.40	0.06	0.05	0.06	0.05	0.06	0.05	0.06	0.05
	0.60	0.06	0.05	0.06	0.05	0.06	0.05	0.05	0.05
	0.70	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.30	0.10	0.05	0.05	0.06	0.04	0.06	0.05	0.06	0.05
	0.40	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.05
	0.60	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05

GV: Generalized variable; W: Williams (1959b) test; OF: Olkin–Finn (1990) test; MRR: Meng, Rosenthal and Rubin (1992) test.

Table 6
(a) Correlations among measures for 66 adult black women and (b) p-values and 90% lower limits for comparing overlapping dependent correlations in (a)

Measure	1	2	3	4
(a)				
1. Body mass index	1.000			
2. Heart rate	0.179	1.000		
3. Systolic blood pressure	0.396	0.088	0.000	
4. Diastolic blood pressure	0.080	−0.042	0.719	1.000

Method	p-values			Lower limits		
	(1)	(2)	(3)	$\rho_{21} - \rho_{41}$	$\rho_{31} - \rho_{21}$	$\rho_{31} - \rho_{41}$
(b)						
GV	0.286	0.082	0.000	−0.128	0.019	0.200
W	0.290	0.084	0.000	−0.131	0.016	0.211
OF	0.286	0.080	0.000	−0.125	0.020	0.204
MRR	0.291	0.086	0.000	–	–	–

(1) $H_a : \rho_{21} > \rho_{41}$; (2) $H_a : \rho_{31} > \rho_{21}$; (3) $H_a : \rho_{31} > \rho_{41}$.

6. Comparison of two non-overlapping dependent correlations

Let ρ_{ij} and ρ_{kl} $(i \neq j \neq k \neq l)$ be any two non-overlapping correlations in a 4×4 normal correlation matrix. Notice that these correlations are non-overlapping because there is no common variable involved. In the following, we provide two asymptotic tests and the GV test.

Table 7
Sizes of the tests for comparing two non-overlapping dependent correlations $H_0 : \rho_{21} \leqslant \rho_{43}$ vs. $H_a : \rho_{21} > \rho_{43}$; $\alpha = 0.05$

$(\rho_{21} = \rho_{43}, \rho_{31}, \rho_{32}, \rho_{41}, \rho_{42})$	$n = 10$			$n = 30$		
	GV	PF	ZPF	GV	PF	ZPF
(0.1, 0.2, 0.6, 0.4, 0.5)	0.05	0.09	0.05	0.05	0.06	0.05
(0.1, 0.2, −0.6, −0.5, 0.2)	0.03	0.09	0.05	0.04	0.06	0.05
(0.1, 0.2, −0.6, −0.1, 0.4)	0.02	0.09	0.05	0.04	0.06	0.05
(0.3, 0.4, −0.3, 0.4, 0.2)	0.05	0.09	0.05	0.06	0.06	0.05
(0.3, 0.4, −0.3, 0.5, 0.1)	0.05	0.10	0.06	0.06	0.06	0.05
(0.3, 0.4, −0.3, −0.2, 0.2)	0.03	0.08	0.05	0.03	0.06	0.05
(0.8, 0.5, 0.6, 0.2, 0.2)	0.03	0.03	0.04	0.04	0.04	0.05
(0.8, 0.5, 0.6, 0.5, 0.4)	0.04	0.03	0.05	0.04	0.04	0.05
$(\rho_{21} = \rho_{43}, \rho_{31}, \rho_{32}, \rho_{41}, \rho_{42})$	$n = 50$			$n = 100$		
	GV	PF	ZPF	GV	PF	ZPF
(0.1, 0.2, 0.6, 0.4, 0.5)	0.05	0.06	0.05	0.05	0.05	0.05
(0.1, 0.2, −0.6, −0.5, 0.2)	0.04	0.06	0.05	0.04	0.05	0.05
(0.1, 0.2, −0.6, −0.1, 0.4)	0.04	0.05	0.05	0.04	0.05	0.05
(0.3, 0.4, −0.3, 0.4, 0.2)	0.05	0.06	0.05	0.05	0.06	0.05
(0.3, 0.4, −0.3, 0.5, 0.1)	0.05	0.06	0.05	0.05	0.06	0.05
(0.3, 0.4, −0.3, −0.2, 0.2)	0.04	0.05	0.05	0.05	0.06	0.05
(0.8, 0.5, 0.6, 0.2, 0.2)	0.05	0.04	0.05	0.04	0.05	0.05
(0.8, 0.5, 0.6, 0.5, 0.4)	0.05	0.05	0.05	0.05	0.05	0.05
(0.8, 0.5, 0.6, −0.1, 0.2)	0.04	0.04	0.05	0.04	0.05	0.05

GV: Generalized variable; PF: Pearson–Filon (1898) method; ZPF: PF based on z transformation.

6.1. *PF (1898) method*

PF test for comparing the two non-overlapping correlations is based on the result that

$$\frac{\sqrt{n}(R_{ij} - R_{kl})}{\sqrt{\text{var}(R_{ij} - R_{kl})}} \sim \text{N}(0, 1) \quad \text{asymptotically}, \tag{24}$$

where the large-sample variance $\text{var}(R_{ij} - R_{kl}) = \text{var}(R_{ij}) + \text{var}(R_{kl}) - 2\,\text{cov}(R_{ij}, R_{kl})$ with $\text{var}(R_{ij}) = (1 - R_{ij}^2)^2$, $\text{var}(R_{kl}) = (1 - R_{kl}^2)^2$ and

$$\begin{aligned}\text{cov}(R_{ij}, R_{kl}) = \tfrac{1}{2}[&(R_{ik} - R_{kj}R_{ij})(R_{lj} - R_{kj}R_{kl}) + (R_{il} - R_{ik}R_{kl})(R_{kj} - R_{ik}R_{ij}) \\ &+ (R_{ik} - R_{il}R_{kl})(R_{lj} - R_{il}R_{ij}) + (R_{il} - R_{ij}R_{lj})(R_{kj} - R_{lj}R_{kl})].\end{aligned} \tag{25}$$

6.2. *The ZPF test*

Another test statistic proposed in the literature (see Raghunathan et al., 1996) is based on Fisher's z transformation, similar to the one given in Section 5.3 for comparing the overlapping dependent correlations. In particular, the statistic

$$\frac{(Z_{ij} - Z_{kl})\sqrt{(n-3)/2}}{1 - \text{cov}(R_{ij}, R_{kl})/(1 - R_{ij}^2)(1 - R_{kl}^2)} \sim \text{N}(0, 1) \quad \text{asymptotically},$$

where Z_{ij} is Fisher's z transformation of R_{ij}, and the large-sample covariance is given in (25).

6.3. *The GV method*

The generalized pivot variable for $\rho_{ij} - \rho_{kl}$ is given by $G_{\rho_{ij}} - G_{\rho_{kl}}$, where $G_{\rho_{uv}}$ is given in (7). The generalized p-value for testing $H_0 : \rho_{ij} \leqslant \rho_{kl}$ vs. $H_a : \rho_{ij} > \rho_{kl}$ is $P(G_{\rho_{ij}} \leqslant G_{\rho_{kl}})$ which can be estimated by Monte Carlo simulation,

Table 8
(a) Correlations of BMI and SBP for three age cohorts; $n = 66$ and (b) p-values and 90% lower limits for comparing non-overlapping dependent correlations in (a)

	Measures	Child		Older sibling		Mother	
		1 BMI	2 SBP	3 BMI	4 SBP	5 BMI	6 SBP
(a)							
Child	1. BMI	1.00					
	2. SBP	0.189	1.00				
Older sibling	3. BMI	0.462	−0.103	1.00			
	4. SBP	0.024	0.013	0.305	1.00		
Mother	5. BMI	0.208	0.143	0.395	0.149	1.00	
	6. SBP	0.023	0.423	−0.091	0.098	0.396	1.00

(b) Method	p-values			Lower limits		
	(1)	(2)	(3)	$\rho_{43} - \rho_{21}$	$\rho_{65} - \rho_{21}$	$\rho_{65} - \rho_{43}$
(b)						
GV	0.240	0.092	0.276	−0.096	0.007	−0.108
PF	0.230	0.088	0.267	−0.090	0.014	−0.101
ZPF	0.241	0.090	0.277	–	–	–

(1) $H_a : \rho_{43} > \rho_{21}$; (2) $H_a : \rho_{65} > \rho_{21}$; (3) $H_a : \rho_{65} > \rho_{43}$.

and the null hypothesis will be rejected whenever the generalized p-value is less than the nominal level α. Furthermore, the percentiles of the generalized pivot variable are used to form a confidence interval for $\rho_{ij} - \rho_{kl}$.

6.4. Simulation studies for non-overlapping dependent correlations

We estimated the sizes of the PF, ZPF and GV tests for $n = 10, 30, 50$ and 100, and present them in Table 7. As noted in Raghunathan et al. (1996), the ZPF test offers improvement over the PF test. The ZPF test, in general, controls the sizes very satisfactorily compared with other two tests, and so it is preferable to the other two tests for practical applications. Even though the GV test is not so accurate, it is useful to obtain confidence intervals for the difference between the correlations being compared.

6.5. An illustrative example for comparing two non-overlapping dependent correlations

Example 1.3 (*Continued*). The correlations between SBP and BMI for the three age cohorts (mother, child and older sibling) and other correlations are given Table 8. We are interested in comparing correlations between SBP and BMI for mother (ρ_{65}), older sibling (ρ_{43}) and child (ρ_{21}). The p-values for comparing correlations in each of the three pairs and 90% lower limits are given in Table 8. As the sample size is large, the results based on all the methods are not appreciably different. However, we note that the p-values of the GV test are closer to those of the ZPF test than the p-values of the PF test.

7. Concluding remarks

In this article, we proposed the GV approach for making inferences about a simple correlation and for comparing two correlations under three different models. We also, via simulation studies, identified the best of the procedures for each problem. Even though the proposed GV approach did not produce superior results than the existing methods, it is conceptually simple and versatile. As mentioned earlier in Section 2, the GV test and confidence limits can be obtained for any scalar-valued function of a normal covariance matrix $\boldsymbol{\Sigma}$. For example, if a dependent variable Y has two sets of

predictors, then the better of the two sets can be determined by comparing the squared multiple correlation coefficients. As the squared multiple correlation coefficient $\mathbf{R}^2$ is a function of $\boldsymbol{\Sigma}$, a generalized pivot variable for $\mathbf{R}^2$ can be easily obtained.

Other problems where the GV tests can be easily obtained include (i) inference on partial correlations, (ii) inference on the generalized variance (determinant of $\boldsymbol{\Sigma}$), (iii) inference on tr($\boldsymbol{\Sigma}$) and (iv) inference on the eigenvalues of $\boldsymbol{\Sigma}$. We are currently working on some of these problems, and plan to publish our work in the literature.

Acknowledgments

The authors are grateful to two reviewers and an associate editor for providing valuable comments and suggestions.

Appendix A.

As pointed out by Weerahandi (1993, p. 900), the problem of finding an appropriate generalized pivotal quantity is a non-trivial task. There is no systematic approach that can be used to find pivotal quantities for all problems. Sufficiency and invariance consideration may reduce the problem. Note that the generalized pivot variable given for ρ_{ij} in (5) is based on the complete sufficient statistic $\boldsymbol{S}$ and is invariant under the transformation $\boldsymbol{S} \to C\boldsymbol{S}C'$, where C is a diagonal matrix with positive diagonal elements.

In general, a generalized pivot variable should satisfy two properties. We shall explain and verify these requirements for $G_{\rho_{ij}}$.

(i) The value of $G_{\rho_{ij}}$ at $\boldsymbol{T}_s = \boldsymbol{t}_s$ should be the parameter of interest. This is true because we see from (4) and (2) that the value of $G_{\boldsymbol{\theta}} = \boldsymbol{t}_s\boldsymbol{V}^{-1} = \boldsymbol{t}_s\boldsymbol{T}_s^{-1}\boldsymbol{\theta}$ at $\boldsymbol{T}_s = \boldsymbol{t}_s$ is $\boldsymbol{\theta}$, and so the value of $G_{\rho_{ij}}$ at $\boldsymbol{T}_s = \boldsymbol{t}_s$ is ρ_{ij}.
(ii) For a given $\boldsymbol{t}_s$, the distribution of $G_{\rho_{ij}}$ should be independent of any unknown parameters. This property also holds because we see from (3) and (4) that the distribution of $G_{\boldsymbol{\theta}}$, when $\boldsymbol{t}_s$ is fixed, does not depend on any parameter.

The generalized test variable $G^{\mathrm{t}}_{\rho_{ij}}$ in (8) should satisfy the following three properties:

(i) The value of $G^{\mathrm{t}}_{\rho_{ij}}$ at $\boldsymbol{T}_s = \boldsymbol{t}_s$ should not depend on any parameter. Here, $G^{\mathrm{t}}_{\rho_{ij}}$ is zero because $G_{\rho_{ij}}$ at $\boldsymbol{T}_s = \boldsymbol{t}_s$ is ρ_{ij}.
(ii) For a given $\boldsymbol{t}_s$, the distribution of $G^{\mathrm{t}}_{\rho_{ij}}$ should depend only on the parameter of interest. Using (8), it is easy to see that this property also holds.
(iii) For a given $\boldsymbol{t}_s$, the distribution of $G^{\mathrm{t}}_{\rho_{ij}}$ should be stochastically monotone with respect to the parameter of interest. From the definition of $G^{\mathrm{t}}_{\rho_{ij}}$ in (8), we see that the distribution of $G^{\mathrm{t}}_{\rho_{ij}}$ is stochastically decreasing with respect to ρ_{ij}.

Thus, we showed that $G_{\rho_{ij}}$ is a bonafide generalized pivot variable for constructing confidence limits for ρ_{ij}, and $G^{\mathrm{t}}_{\rho_{ij}}$ is a valid generalized test variable for hypothesis testing about ρ_{ij}. For example, the 100αth percentile of $G_{\rho_{ij}}$ is a $100(1-\alpha)\%$ lower confidence limit for ρ_{ij}. If one is interested in testing

$$\mathrm{H}_0 : \rho_{ij} \leqslant \rho_{ij,0} \text{ vs. } \mathrm{H}_a : \rho_{ij} > \rho_{ij,0}, \tag{A.1}$$

where $\rho_{ij,0}$ is a specified number, then, noting that $G^{\mathrm{t}}_{\rho_{ij}}$ is stochastically decreasing in ρ_{ij}, the generalized p-value for (A.1) is given by

$$\sup_{\mathrm{H}_0} P(G^{\mathrm{t}}_{\rho_{ij}} \leqslant 0) = P(G^{\mathrm{t}}_{\rho_{ij}} \leqslant 0 | \rho_{ij} = \rho_{ij,0}) = P(G_{\rho_{ij}} \leqslant \rho_{ij,0}). \tag{A.2}$$

Appendix B.

Let $\boldsymbol{s}$ and $\boldsymbol{r}$ be the observed values for $\boldsymbol{S}$ and $\boldsymbol{R}$, respectively. Let $\boldsymbol{D}_s = \mathrm{diag}(\sqrt{s_{11}}, \ldots, \sqrt{s_{pp}})$. Then we have

$$\boldsymbol{r} = \boldsymbol{D}_s^{-1}\boldsymbol{s}\boldsymbol{D}_s^{-1}.$$

Let $\boldsymbol{t}_s = (t_{s,ij})$ and $\boldsymbol{t}_r = (t_{r,ij})$ be the Cholesky decompositions of $\boldsymbol{s}$ and $\boldsymbol{r}$, respectively. That is, $\boldsymbol{s} = \boldsymbol{t}_s\boldsymbol{t}_s'$ and $\boldsymbol{r} = \boldsymbol{t}_r\boldsymbol{t}_r'$, where $\boldsymbol{t}_s$ and $\boldsymbol{t}_r$ are lower triangular matrices with positive diagonal elements. It is easy to see that

$$t_{r,11} = 1, \quad s_{kk} = \sum_{j=1}^{k} t_{s,kj}^2, \quad k = 1, \ldots, p.$$

Since $\boldsymbol{t}_r\boldsymbol{t}_r' = \boldsymbol{r} = \boldsymbol{D}_s^{-1}\boldsymbol{s}\boldsymbol{D}_s^{-1} = \boldsymbol{D}_s^{-1}\boldsymbol{t}_s\boldsymbol{t}_s'\boldsymbol{D}_s^{-1}$, we have $\boldsymbol{t}_r = \boldsymbol{D}_s^{-1}\boldsymbol{t}_s$ or equivalently, $t_{r,ij} = t_{s,ij}/\sqrt{s_{ii}}$ for $i \geqslant j$. This implies that

$$\frac{t_{r,ij}}{t_{r,ii}} = \frac{t_{s,ij}}{t_{s,ii}}, \quad i \geqslant j. \tag{B.1}$$

Let $\boldsymbol{A} = \boldsymbol{t}_s\boldsymbol{V}^{-1} = (a_{ij})$ $\boldsymbol{B} = \boldsymbol{t}_r\boldsymbol{V}^{-1} = (b_{ij})$ and $\boldsymbol{V}^{-1} = (v^{ij})$. Noticing that all the matrices are lower triangular, we have

$$a_{ij} = \sum_{l=j}^{i} t_{s,il}v^{lj} \quad \text{and} \quad b_{ij} = \sum_{l=j}^{i} t_{r,il}v^{lj}, \quad i \geqslant j. \tag{B.2}$$

It follows from (B.2) and (B.1) that

$$\frac{a_{ij}}{a_{ii}} = \frac{\sum_{l=j}^{i} t_{s,il}v^{lj}}{t_{s,ii}v^{ii}} = \sum_{l=j}^{i}\left(\frac{t_{s,il}}{t_{s,ii}}\right)\left(\frac{v^{il}}{v^{ii}}\right) = \sum_{l=j}^{i}\left(\frac{t_{r,il}}{t_{r,ii}}\right)\left(\frac{v^{il}}{v^{ii}}\right) = \frac{b_{ij}}{b_{ii}} \quad \text{for } i > j. \tag{B.3}$$

Thus, the GV for ρ_{ij} in (6) can be written as

$$G_{\rho_{ij}} = \frac{\sum_{k=1}^{j}(a_{ik}/a_{ii})(a_{jk}/a_{jj})}{\sqrt{1+\sum_{k=1}^{i-1}(a_{ik}/a_{ii})^2}\sqrt{1+\sum_{k=1}^{j-1}(a_{jk}/a_{jj})^2}} \quad \text{for } i > j. \tag{B.4}$$

Using (B.3) in (B.4), and after some simplification, we get (7).

References

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Bilker, W.B., Brensinger, C., Gur, R.C., 2004. A two factor ANOVA-like test for correlated correlations: CORANOVA. Multivariate Behavioral Res. 39, 565–594.

Bland, J.M., Altman, D.G., 1996. Statistics notes: measurement error and correlation coefficients. British Medical J. 313, 41–42.

Casella, G., Berger, R.L., 2002. Statistical Inference. Duxbury, Pacific Grove, CA.

Fisher, L.D., Van Belle, G., 1993. A Methodology for the Health Sciences. Wiley, New York.

Fisher, R.A., 1921. On the probable error of a coefficient of correlation deduced from a small sample. Metron 1, 3–32.

Gamage, J., Mathew, T., Weerahandi, S., 2004. Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem. J. Multivariate Anal. 88, 177–189.

Hotelling, H., 1940. The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Ann. Math. Statist. 11, 271–283.

Hotelling, H., 1953. New light on the correlation coefficient and its transform. J. Roy. Statist. Soc. Ser. B 15, 193–225.

Jayaratnam, S., 1992. Confidence intervals for the correlation coefficient. Statist. Probab. Lett. 15, 389–393.

Krishnamoorthy, K., Lu, Y., 2003. Inferences on the common mean of several normal populations based on the generalized variable method. Biometrics 59, 237–247.

Krishnamoorthy, K., Mathew, T., 2003. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. J. Statist. Plann. Inference 115, 103–121.

Krishnamoorthy, K., Mathew, T., 2004. One-Sided tolerance limits in balanced and unbalanced one-way random models based on generalized confidence limits. Technometrics 46, 44–52.

Meng, X.L., Rosenthal, R., Rubin, D.B., 1992. Comparing correlated correlation coefficients. Psych. Bull. 111, 172–175.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

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

Olkin, I., Finn, J.D., 1990. Testing correlated correlations. Psych. Bull. 108, 330–333.

Olkin, I., Finn, J.D., 1995. Correlations redux. Psych. Bull. 118, 155–164.

Pearson, K., Filon, L.N.G., 1898. Mathematical contributions to the theory of evolution. Trans. Roy. Soc. London (Ser. A) 91, 259–262.

Raghunathan, T.E., Rosenthal, R., Rubin, D.B., 1996. Comparing correlated but nonoverlapping correlations. Psych. Methods 1, 178–183.

Tsui, K.W., Weerahandi, S., 1989. Generalized *p*-value in significance testing of hypotheses in the presence of nuisance parameters. J. Amer. Statist. Assoc. 84, 602–607.
Weerahandi, S., 1993. Generalized confidence intervals. J. Amer. Statist. Assoc. 88, 899–905.
Weerahandi, S., 1995a. Exact Statistical Methods for Data Analysis. Springer, New York.
Weerahandi, S., 1995b. ANOVA under unequal error variances. Biometrics 51, 589–599.
Weerahandi, S., Berger, V.W., 1999. Exact inference for growth curves with intraclass correlation structure. Biometrics 55, 921–924.
Williams, E.J., 1959a. Significance of difference between two non-independent correlation coefficients. Biometrics 15, 135–136.
Williams, E.J., 1959b. The comparison of regression variables. J. Roy. Statist. Soc. Ser. B 21, 396–399.
Zhou, L., Mathew, T., 1994. Some tests for variance components using generalized *p*-values. Technometrics 36, 394–402.