Psychological Bulletin
1980, Vol. 87, No. 2, 245–251

Tests for Comparing Elements of a Correlation Matrix

James H. Steiger
University of British Columbia, Vancouver, British Columbia, Canada

In a variety of situations in psychological research, it is desirable to be able to make statistical comparisons between correlation coefficients measured on the same individuals. For example, an experimenter may wish to assess whether two predictors correlate equally with a criterion variable. In another situation, the experimenter may wish to test the hypothesis that an entire matrix of correlations has remained stable over time. The present article reviews the literature on such tests, points out some statistics that should be avoided, and presents a variety of techniques that can be used safely with medium to large samples. Several illustrative numerical examples are provided.

Statistical tests for comparing two or more elements of a correlation matrix are of considerable potential value for interpreting the outcomes of psychological research. As a simple example, suppose an experimenter wishes to test the hypothesis that a predictor-criterion correlation has not changed from an earlier value obtained on the same subjects. This can be treated statistically as a hypothesis of the form $\rho_{12} = \rho_{34}$. Unfortunately, the test of such a hypothesis is not straightforward because correlation coefficients measured on the same individuals are not, in general, independent. However, in the past 10 years, mathematical statisticians have developed efficient methods for testing such a hypothesis. Regrettably, the better techniques have not yet filtered through to psychological statistics texts, and some authoritative sources cite methods that are clearly suboptimal.

The purpose of the present article is to review the recent literature and correct some misconceptions about tests for comparing correlation coefficients. In addition, some new techniques are presented that yield improved small-sample performance and computational efficiency. Several techniques are illustrated with numerical examples.

The author gratefully acknowledges the research support of National Research Council of Canada Grant A4640 and University of British Columbia Humanities and Social Sciences Grant 26-9914.

Requests for reprints should be sent to James H. Steiger, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5.

Tests for the Equality of Two Dependent Correlations

Two sample correlations obtained on the same individuals are not, in general, independent of each other. Indeed, correlations have a correlation matrix of their own. As early as 1898, Pearson and Filon obtained asymptotic expressions for the variance–covariance matrix of a set of correlations. Denoting the covariance between r_{jk} and r_{jh} as $\sigma_{jk,jh}$ and the variance of r_{jk} as σ_{jk}^2, expressions equivalent to those of Pearson and Filon are

$$\psi_{jk}^2 = N\sigma_{jk}^2 = (1 - \rho_{jk}^2)^2; \tag{1}$$

and

$$\begin{aligned}\psi_{jk,hm} = N\sigma_{jk,hm} = \tfrac{1}{2}\{&[(\rho_{jh} - \rho_{jk}\rho_{kh}) \\ &\times(\rho_{km} - \rho_{kh}\rho_{hm})] + [(\rho_{jm} - \rho_{jh}\rho_{hm}) \\ &\times(\rho_{kh} - \rho_{kj}\rho_{jh})] + [(\rho_{jh} - \rho_{jm}\rho_{mh}) \\ &\times(\rho_{km} - \rho_{kj}\rho_{jm})] + [(\rho_{jm} - \rho_{jk}\rho_{km}) \\ &\times(\rho_{kh} - \rho_{km}\rho_{mh})]\}.\end{aligned} \tag{2}$$

When two correlations have an index in common, the covariance expression in Equation 2 can be simplified to the equivalent form,

$$\begin{aligned}\psi_{jk,jh} = N\sigma_{jk,jh} = {}&\rho_{kh}(1 - \rho_{jk}^2 - \rho_{jh}^2) \\ &-\tfrac{1}{2}(\rho_{jk}\rho_{jh})(1 - \rho_{jk}^2 - \rho_{jh}^2 - \rho_{kh}^2).\end{aligned} \tag{3}$$

The above results led to some early asymptotic z tests for null hypotheses of the forms $\rho_{jk} = \rho_{hm}$ and $\rho_{jk} = \rho_{jh}$. It is well-known from asymptotic distribution theory that as the sample size becomes large, the distribution of a set of correlation coefficients approaches the multivariate normal in form. Hence, with large samples, two sample correlation coefficients (and their differences) have an approximately normal distribution. To construct a practical test statistic, one must estimate the ψ_{jk}^2 and the $\psi_{jk,hm}$ with some consistent estimator, but this will not affect the asymptotic result. Hence, if $\hat{\psi}$ denotes expressions obtained by substituting r_{jk} for ρ_{jk} in Equations 1–3, then

$$Z_1 = N^{\frac{1}{2}}(r_{jk} - r_{jh}) \times (\hat{\psi}_{jk}^2 + \hat{\psi}_{jh}^2 - 2\hat{\psi}_{jk,jh})^{-\frac{1}{2}} \quad (4)$$

will have an asymptotic distribution that is standard normal. Statistic Z_1 provides a large-sample statistic for testing the equality of two correlations with one index in common. Similarly, the statistic,

$$Z_2 = N^{\frac{1}{2}}(r_{jk} - r_{hm}) \times (\hat{\psi}_{jk}^2 + \hat{\psi}_{hm}^2 - 2\hat{\psi}_{jk,hm})^{-\frac{1}{2}}, \quad (5)$$

allows a large-sample test of the hypothesis $\rho_{jk} = \rho_{hm}$. Z_1 and Z_2 were popularized by Peters and Van Voorhis (1940) in their classic introductory text.

Hotelling (1940) suggested an alternative to Z_1 for testing null hypotheses of the form $\rho_{jk} = \rho_{jh}$. Hotelling's statistic was

$$T_1 = (N - 3)^{\frac{1}{2}}(r_{jk} - r_{jh}) \times (1 + r_{kh})^{\frac{1}{2}}(2|\mathbf{R}|)^{-\frac{1}{2}}, \quad (6)$$

where

$$|\mathbf{R}| = (1 - r_{jk}^2 - r_{jh}^2 - r_{kh}^2) + (2r_{jk}r_{jh}r_{kh})$$

is the determinant of the 3×3 correlation matrix containing the coefficients being tested. Under highly restrictive assumptions, T_1 is distributed as Student's t with $df = N - 3$.

Although Z_1 and Z_2 are useful for large samples, Hotelling's T_1 is basically useless as a replacement for Z_1 with any sample size because it does not have its designated distribution (or even come close to it) under a variety of conditions in which $\rho_{jk} = \rho_{jh}$. For example, if $\rho_{12} = \rho_{13} = (.5)^{\frac{1}{2}}$ and $\rho_{23} = 0$, the null hypothesis is true. Yet, at a nominal Type I error rate (α) of .05, T_1 will almost always reject the null hypothesis. Williams (1959) proposed a modification of T_1 to alleviate this difficulty. Williams's formula is

$$T_2 = (r_{jk} - r_{jh}) \sqrt{\frac{(N-1)(1+r_{kh})}{2\left(\frac{N-1}{N-3}\right)|\mathbf{R}| + \bar{r}^2(1-r_{kh})^3}}, \quad (7)$$

where $\bar{r} = \frac{1}{2}(r_{jk} + r_{jh})$. T_2, like T_1, has a t distribution with $df = N - 3$. Unfortunately, some prominent psychological statistics texts (Ferguson, 1976; McNemar, 1969) and articles (Kenny, 1975) have recommended T_1 for comparing two correlations with an index in common. However, T_1 need not and should not be used for this purpose.

In 1964, Olkin and Siotani (Note 1) developed a number of important results in correlational theory, including a succinct rederivation of Equations 1–3. Olkin (1967) provided a readable, popularized account of several correlational statistics, including Z_1 and Z_2. Unfortunately, formulas for Z_1 and Z_2 given in Olkin's text have typographical errors. Glass and Stanley (1970) give a correct formula for Z_1. Kenny (1973) recommends Z_2 as a test of significance in cross-lagged panel correlation analysis.

The normality of Z_1 and Z_2 depends on the asymptotic normality of sample correlation coefficients. However, if the sample is not large and population correlations have extreme values, these statistics depart from their nominal Type I error rates (Steiger, Note 2). The Fisher (1921) r-to-z transform,

$$z_{jk} = \tfrac{1}{2} \ln\left(\frac{1 + r_{jk}}{1 - r_{jk}}\right), \quad (8)$$

helps to eliminate this problem because it transforms a sample correlation to a variable that is close to normally distributed, even with small sample sizes and extreme ρ_{jk}. To capitalize on the virtues of this transformation, Dunn and Clark (1969) developed statistics that are analogous to Z_1 and Z_2 but that use the Fisher transform instead of raw correlations. As a straightforward consequence of a theorem in Olkin and Siotani (Note 1), it can be shown, the z_{jk} have asymptotic variances and covariances that can be expressed as

$$c_{jk}^2 = (N - 3)\sigma_{z_{jk}}^2 = 1; \quad (9)$$

$$c_{jk,jh} = (N-3)\sigma_{z_{jk},z_{jh}} = \psi_{jk,jh}\psi_{jk}^{-1}\psi_{jh}^{-1} = \psi_{jk,jh}/(1-\rho_{jk}^2)(1-\rho_{jh}^2); \quad (10)$$

and

$$c_{jk,hm} = (N-3)\sigma_{z_{jk},z_{hm}} = \psi_{jk,hm}\psi_{jk}^{-1}\psi_{hm}^{-1} = \psi_{jk,hm}/(1-\rho_{jk}^2)(1-\rho_{hm}^2). \quad (11)$$

In the following discussion, sample estimates of $c_{jk,jh}$ and $c_{jk,hm}$, obtained by substituting sample correlations for population correlations in Equations 10 and 11, are denoted $s_{jk,jh}$ and $s_{jk,hm}$, respectively.

Fisher-transformed correlations have an asymptotic distribution that like that of the r_{jk}, is multivariate normal. However, the z_{jk} retain their marginal normality in small samples, and their variance in small samples is extremely close to the asymptotic value $1/(N-3)$. Hence, we would have strong intuitive reason to suspect that Equations 9–11 could be used to obtain improved analogs of statistics Z_1 and Z_2. These statistics, given in Dunn and Clark (1969), are

$$Z_1^* = (N-3)^{\frac{1}{2}}(z_{jk}-z_{jh})(2-2s_{jk,jh})^{-\frac{1}{2}}; \quad (12)$$

and

$$Z_2^* = (N-3)^{\frac{1}{2}}(z_{jk}-z_{hm})(2-2s_{jk,hm})^{-\frac{1}{2}}. \quad (13)$$

Monte Carlo simulation experiments (Neill & Dunn, 1975; Steiger, Note 2) have confirmed that statistics Z_1^* and Z_2^* (as well as T_2) are notably superior to Z_1 and Z_2 in maintaining Type I error rate control at small sample sizes. All three statistics can be used with confidence when sample size exceeds 20.

Further improvement in Type I error rate control can be obtained by modifying Z_1^* and Z_2^* to incorporate the null hypothesis. Specifically, we estimate those correlations that are equal under the null hypothesis by pooling corresponding sample correlations. For example, if the null hypothesis $\rho_{jk} = \rho_{hm}$ is true, then $\bar{r}_{jk,hm} = \frac{1}{2}(r_{jk} + r_{hm})$ gives a pooled, more reliable estimate (an ordinary least squares estimate) of both ρ_{jk} and ρ_{hm} than either sample correlation taken separately. The pooled estimate can be used in place of r_{jk} and r_{hm} in Equation 11 for computing $s_{jk,hm}$. If we denote an $s_{jk,hm}$ computed with a pooled estimate as $\bar{s}_{jk,hm}$, then the modified test statistics can be written

$$\bar{Z}_1^* = (N-3)^{\frac{1}{2}}(z_{jk}-z_{jh})(2-2\bar{s}_{jk,jh})^{-\frac{1}{2}}; \quad (14)$$

and

$$\bar{Z}_2^* = (N-3)^{\frac{1}{2}}(z_{jk}-z_{hm})(2-2\bar{s}_{jk,hm})^{-\frac{1}{2}}. \quad (15)$$

Comparisons Among More Than Two Correlations

A large number of hypotheses involving the comparison of two correlations can be subsumed under a common descriptive heading as pattern hypotheses and can be tested statistically within the same general framework. A pattern hypothesis on a population correlation matrix $\mathbf{P}$ is any hypothesis that states that some of its elements are equal to each other and/or to specified numerical values.

If we extract the $k = (m^2 - m)/2$ unique off-diagonal elements of $\mathbf{P}$, the $m \times m$ population correlation matrix, and place them in a $k \times 1$ vector $\mathbf{p}$, we can express any pattern hypothesis in the form

$$H_0\colon \mathbf{p} = \mathbf{p}_0 = \Delta\gamma + \mathbf{p}^*, \quad (16)$$

where Δ is a $k \times q$ matrix of zeros and ones with elements $\delta_{ij} = \partial \mathbf{p}_{0i}/\partial\gamma_j$, γ is a $q \times 1$ vector of common (unspecified) correlations, and $\mathbf{p}^*$ is a $k \times 1$ vector containing specified values for elements of $\mathbf{p}$ and containing zeros in other positions. For example, let $\mathbf{P}$ be 4×4. Let H_0 be that $\rho_{31} = \rho_{21} = \gamma_1$, that $\rho_{32} = .6$ (a specified value), and that $\rho_{41} = \rho_{42} = \rho_{43} = \gamma_2$. Then H_0 may be written in the matrix notation of Equation 16 as

$$\begin{bmatrix}\rho_{21}\\ \rho_{31}\\ \rho_{32}\\ \rho_{41}\\ \rho_{42}\\ \rho_{43}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 1 & 0\\ 0 & 0\\ 0 & 1\\ 0 & 1\\ 0 & 1\end{bmatrix} \begin{bmatrix}\gamma_1\\ \gamma_2\end{bmatrix} + \begin{bmatrix}0\\ 0\\ .6\\ 0\\ 0\\ 0\end{bmatrix}$$
$$\mathbf{p} \quad = \quad \Delta \quad \gamma \quad + \quad \mathbf{p}^* .$$

There are a number of alternative approaches for estimation and hypothesis testing of correlational pattern hypotheses. Estimates of $\mathbf{p}_0$ generally take the form $\hat{\mathbf{p}} = \Delta\hat{\gamma} + \mathbf{p}^*$, so the problem of estimating $\hat{\mathbf{p}}$ is essentially equivalent to obtaining an estimate $(\hat{\gamma})$ for γ.

Maximum likelihood estimates $(\hat{\mathbf{p}}_{\mathrm{ML}})$ can be obtained by treating the pattern hypothesis as a special case of the analysis of covariance structures. McDonald (1974, 1975) developed

TESPAR, a highly efficient computer algorithm that yields maximum likelihood estimates for the elements of $\mathbf{p}$ (and the population covariance matrix) when $\mathbf{p}^* = \emptyset$, a null vector. These estimates can then be used to calculate a likelihood ratio chi-square statistic with $df = k - q$. The statistic is of the form $U = N\Phi$, with

$$\Phi = ln|\mathbf{C}^*| - ln|\mathbf{S}| + \text{trace}\,(\mathbf{SC}^{*-1}) - m, \tag{17}$$

where $\mathbf{C}^*$ is the maximum likelihood estimate of $\mathbf{C}$, the population covariance matrix, under H_0, and $\mathbf{S}$ is the sample covariance matrix.

The maximum likelihood (likelihood ratio test) approach, as currently used, appears to have at least two major drawbacks. First, it requires sometimes lengthy computer iteration of $\hat{\mathbf{p}}_{ML}$. Second, U is a large-sample statistic that can be shown, through Monte Carlo simulations, to reject the null hypothesis too frequently for moderate to small sample sizes (see Neill & Dunn, 1975; Steiger, Note 3).

The latter difficulty can be overcome, for all practical purposes, by using $\hat{\mathbf{p}}_{ML}$ in a quadratic form statistic that like the Dunn-Clark and Neill-Dunn statistics, uses the distributional stability of the Fisher transform to achieve improved small-sample performance. Let $\mathbf{r}$ be a vector of observed sample correlations, with indices corresponding to the elements of $\mathbf{p}$. Let $\mathbf{z}(\hat{\mathbf{p}}_{ML})$ be a vector of Fisher transforms of the elements of $\hat{\mathbf{p}}_{ML}$, and define $\mathbf{z}(\mathbf{r})$ likewise for $\mathbf{r}$. Define $\mathbf{S}_{ML}$ as N times the estimated variance–covariance matrix for $\mathbf{z}(\mathbf{r})$. Elements of $\mathbf{S}_{ML}$ are obtained by substituting elements of $\hat{\mathbf{p}}_{ML}$ for population correlations in Equations 9–11. Then the statistic,

$$X_1 = (N-3)[\mathbf{z}(\mathbf{r}) - \mathbf{z}(\hat{\mathbf{p}}_{ML})]' \times \mathbf{S}_{ML}^{-1}[\mathbf{z}(\mathbf{r}) - \mathbf{z}(\hat{\mathbf{p}}_{ML})], \tag{18}$$

has, like U, an asymptotic χ^2_{k-q} distribution.

An alternative approach to estimation and testing of pattern hypotheses, one that avoids lengthy computer iteration, is based on the use of generalized least squares estimators ($\hat{\mathbf{p}}_{GLS}$). The calculation of $\hat{\mathbf{p}}_{GLS}$ proceeds as follows: Define $\hat{\mathbf{p}}_{LS}$ as a vector of ordinary least squares estimators for $\mathbf{p}_0$. As a straightforward extension of the estimator $\bar{\mathbf{r}}_{jk,hm}$ discussed earlier, $\hat{\mathbf{p}}_{LS}$ is given as

$$\hat{\mathbf{p}}_{LS} = \Delta\hat{\gamma}_{LS} + \mathbf{p}^*; \quad \hat{\gamma}_{LS} = (\Delta'\Delta)^{-1}\Delta'(\mathbf{r} - \mathbf{p}^*). \tag{19}$$

Next, we define an estimate of N times the variance–covariance matrix of $\mathbf{r}$, $\hat{\Sigma}_{LS}$, as the matrix whose elements are obtained by substituting elements of $\hat{\mathbf{p}}_{LS}$ for ρ_{jk} in Equations 1–3. Then $\hat{\mathbf{p}}_{GLS}$, the vector of generalized least squares estimates, is

$$\hat{\mathbf{p}}_{GLS} = \Delta\hat{\gamma}_{GLS} + \mathbf{p}^*;$$
$$\hat{\gamma}_{GLS} = (\Delta'\hat{\Sigma}_{LS}{}^{-1}\Delta)^{-1}[\Delta'\hat{\Sigma}_{LS}{}^{-1}(\mathbf{r} - \mathbf{p}^*)]. \tag{20}$$

Browne (1974) showed that $\hat{\mathbf{p}}_{GLS}$ is asymptotically equivalent to $\hat{\mathbf{p}}_{ML}$. If we define $\mathbf{S}_{LS}$ as the matrix whose elements are obtained by substituting elements of $\hat{\mathbf{p}}_{LS}$ for ρ_{jk} in Equations 9–11, then

$$X_2 = (N-3)[\mathbf{z}(\mathbf{r}) - \mathbf{z}(\hat{\mathbf{p}}_{GLS})]' \times \mathbf{S}_{LS}^{-1}[\mathbf{z}(\mathbf{r}) - \mathbf{z}(\hat{\mathbf{p}}_{GLS})] \tag{21}$$

is, like X_1, asymptotically χ^2_{k-q}. Browne (1977), who has pioneered the use of generalized least squares estimators for testing pattern hypotheses, has pointed out that in many cases, $\hat{\mathbf{p}}_{GLS}$ and $\hat{\mathbf{p}}_{LS}$ are formally equivalent and that, generally, they differ only slightly. Hence, in situations in which computational convenience is paramount, an approximate χ^2_{k-q} statistic may be computed by substituting $\hat{\mathbf{p}}_{LS}$ for $\hat{\mathbf{p}}_{GLS}$ in Equation 21. Recent unpublished Monte Carlo experiments (Steiger, Note 3) on the relative performance of U, X_1, and X_2 indicate that U is notably inferior in Type I error rate (α) control to X_1 and X_2 at small to moderate sample sizes. U rejects the null hypothesis too often, whereas X_1 and X_2 have essentially equivalent performance. The latter two statistics maintain essentially nominal α for $N > 50$, and performance is adequate for samples as small as 20.

Some Numerical Examples

The following numerical examples provide concrete illustrations of many of the computational methods. To consolidate the various examples, all significance tests ($\alpha = .05$) are based on the same hypothetical correlation matrix, given in Table 1. Table 1 contains sample correlations, based on 103 observations, for a hypothetical longitudinal study of sex stereotypes and verbal achievement. Three variables are measured twice on the same individuals.

Case A: H_0: $\rho_{jk} = \rho_{jh}$

A number of efficient statistics are available for testing this hypothesis. Computations for T_2 and $\bar{Z}_1^*$ are illustrated. Suppose that the experimenter hypothesizes that $\rho_{31} = \rho_{32}$, that is, that masculinity and femininity scores correlate equally with verbal achievement at Time 1. We have $r_{31} = .40$, $r_{32} = .50$, $r_{12} = .10$, $N = 103$, $|\mathbf{R}| = .62$, and $\bar{r} = \frac{1}{2}(r_{31} + r_{32}) = .45$. Substituting in Equation 7, we obtain $T_2 = -.8913$, which, referred to the t distribution with $df = 100$, is not significant. To compute $\bar{Z}_1^*$, we substitute $\bar{r} = .45$ for r_{31} and r_{32} in computing $s_{31,32}$. Hence, from Equations 3 and 10,

$$s_{31,32} = [(.10)(1 - .45^2 - .45^2) - \tfrac{1}{2}(.45)(.45)(1 - .45^2 - .45^2 - .10^2)]/(1 - .45^2)(1 - .45^2) = .0042.$$

$\bar{Z}_1^*$ is then computed from Equation 14 as $-.8890$. This value, when compared to the standard normal curve rejection points of ± 1.96, is not significant.

Case B: H_0: $\rho_{jk} = \rho_{hm}$

In this case, suppose the experimenter were interested in testing the hypothesis that the correlation between femininity and verbal achievement was the same at Times 1 and 2. This hypothesis, that $\rho_{32} = \rho_{65}$, can be tested with $\bar{Z}_2^*$ as follows. First, $\bar{r}_{32,65} = \frac{1}{2}(.50 + .60) = .55$ is substituted for r_{32} and r_{65} in computing $\bar{s}_{32,65}$ from Equation 11. We also have $r_{jk} = r_{36} = .80$, $r_{kh} = r_{26} = .50$, $r_{km} = r_{25} = .70$, $r_{jm} = r_{35} = .50$, and $N = 103$. Substituting in Equations 11 and 2, we obtain $2\bar{s}_{32,65} = .9517$. Since $z_{32} = .5493$ and $z_{65} = .6931$, $\bar{Z}_2^*$ may be computed from Equation 15 as -1.4045. On the basis of this result, the null hypothesis would not be rejected.

Case C: The Identity Hypothesis, H_0: $\mathbf{P} = \mathbf{I}$

As Larzelere and Mulaik (1977) have pointed out, performance of a large number of individual significance tests of the form $\rho_{ij} = 0$ can be analogous to the fallacy of multiple t tests in the analysis of variance unless special statistical precautions are taken. One way of obtaining experimentwise error-rate protection in such situations is to routinely perform a simultaneous hypothesis test that all correlations are 0 or that $\mathbf{P} = \mathbf{I}$, an identity matrix. If the preliminary overall test fails to reject, then no further individual tests are performed. Larzelere and Mulaik reviewed several methods for testing the identity hypothesis, but the present method, a simple special case of statistic X_2, is computationally much simpler than the tests they mention and, unlike these other tests, can be performed easily by hand. Specifically, if $\mathbf{P} = \mathbf{I}$, then $\mathbf{S}_{LS} = \mathbf{I}$, and $\hat{\boldsymbol{\rho}}_{GLS} = \emptyset$, a null vector. Hence, under this null hypothesis, Equation 21 reduces to

$$X_2 = (N - 3) \sum_{j<i} \sum z_{ij}^2. \qquad (22)$$

Applying the test statistic to the correlation matrix in Table 1, we obtain $X_2 = 543.17$, which, when referred to a χ^2_{15} distribution, leads to overwhelming rejection of the null hypothesis.

Table 1
Correlation Matrix for a Hypothetical Longitudinal Study

Variable	M1	F1	V1	M2	F2	V2
M1	—					
F1	.10	—				
V1	.40	.50	—			
M2	.70	.05	.50	—		
F2	.05	.70	.50	.50	—	
V2	.45	.50	.80	.50	.60	—

Note. M1 = masculinity at Time 1; F1 = femininity at Time 1; V1 = verbal ability at Time 1; M2 = masculinity at Time 2; F2 = femininity at Time 2; V2 = verbal ability at Time 2.

Case D: H_0: $\rho_{21} = \rho_{31} = \rho_{32} = \gamma_1$

In this example, the experimenter hypothesizes that the three variables have equal intercorrelations at Time 1. This hypothesis may be expressed in the notation of Equation 16 as

$$\begin{bmatrix}\rho_{21}\\ \rho_{31}\\ \rho_{32}\end{bmatrix} = \begin{bmatrix}1\\ 1\\ 1\end{bmatrix} [\gamma_1] + \begin{bmatrix}0\\ 0\\ 0\end{bmatrix}.$$

Table 2
Null Hypothesis of Longitudinal Stability

Elements of p		Elements of Δ	Elements of γ		Elements of $\mathbf{p}^*$

$$
\begin{bmatrix}\rho_{21}\\ \rho_{31}\\ \rho_{32}\\ \rho_{41}\\ \rho_{42}\\ \rho_{43}\\ \rho_{51}\\ \rho_{52}\\ \rho_{53}\\ \rho_{54}\\ \rho_{61}\\ \rho_{62}\\ \rho_{63}\\ \rho_{64}\\ \rho_{65}\end{bmatrix}
=
\begin{bmatrix}
1&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0\\
1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&1\\
0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0
\end{bmatrix}
\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \gamma_4\\ \gamma_5\\ \gamma_6\\ \gamma_7\\ \gamma_8\\ \gamma_9\\ \gamma_{10}\\ \gamma_{11}\\ \gamma_{12}\end{bmatrix}
+
\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\end{bmatrix}
$$

We find in this case that $\hat{\gamma}_{GLS} = \hat{\gamma}_{LS} = \frac{1}{3}$. Using Equation 21, we obtain $X_2 = 14.95$. In this case, $df = k - q = 3 - 1 = 2$, and H_0 is rejected.

Case E: Testing Longitudinal Stability of a Correlation Matrix

A potentially important use for the pattern hypothesis is in testing whether a set of correlations (or perhaps an entire correlation matrix) has remained constant at two (or more) points in time. In the present example, suppose the experimenter wishes to hypothesize that the intercorrelations among masculinity, femininity, and verbal achievement remain constant at Times 1 and 2. This null hypothesis, which specifies simultaneously that $\rho_{21} = \rho_{54} = \gamma_1$, $\rho_{31} = \rho_{64} = \gamma_2$, and $\rho_{32} = \rho_{65} = \gamma_3$, can be stated in the notation of Equation 16 as shown in Table 2. In this case, X_2 can be computed as $X_2 = 34.097$, which, when referred to the χ^2_3 distribution, is significant.

Conclusions and Recommendations

Tests for Comparing Two Dependent Correlations

Some highly efficient techniques are available, and empirical evidence suggests that they can be used with confidence on sample sizes as small as 20. The decision concerning which of the good techniques to use is perhaps less important than the knowledge of which tests are clearly suboptimal and should be avoided. When the null hypothesis of interest is of the form $\rho_{jk} = \rho_{jh}$, tests T_1 and Z_1 should *not* be used, whereas T_2, Z_1^*, and $\bar{Z}_1^*$ are acceptable, with T_2 perhaps the best all-round choice.

When the null hypothesis is of the form $\rho_{jk} = \rho_{hm}$, tests Z_2^* and $\bar{Z}_2^*$ are acceptable, with the latter probably preferable in most applications as the slightly more conservative statistic. Z_2, on the other hand, should never be used because its performance is markedly inferior to the other two statistics. All three require about the same computational effort.

Tests for Comparing Several Dependent Correlations

Many hypotheses of interest involving the comparison of several dependent correlations can be expressed as pattern hypotheses and can be tested using the general techniques described. The more traditional approach has been to compute $\hat{\mathbf{p}}_{ML}$, followed by the likelihood ratio test statistic U. However, it is now clear from Monte Carlo research that if the maximum likelihood approach is used, U is not an optimal test statistic because it rejects a true null hypothesis too often at small to moderate sample sizes. One solution to this problem that works well is to compute the quadratic form test statistic X_1 instead of U. An alternative approach might be to develop a generalized correction constant to be used in place of N in calculating U.

Unfortunately, even if an efficient statistic is calculated instead of U, $\hat{\rho}_{ML}$ must be obtained by computer iteration. On the other hand, iteration is not required in calculating the generalized least squares estimates. $\hat{\rho}_{GLS}$ shares many of the asymptotic properties of $\hat{\rho}_{ML}$, tends in practice to be almost identical to it, even at small samples, and can be computed at a fraction of the cost (in many cases, if necessary, with an advanced hand calculator). Hence, for most practical purposes, the generalized least squares approach yields hypothesis tests that are, at worst, only slightly less accurate than those obtained by maximum likelihood methods but that require far less computational effort. On balance, the generalized least squares approach has much to recommend it.[1]

[1] A computer program, MULTICORR, that computes statistic X_2 for any pattern hypothesis on correlation matrices of order 20×20 or less is available from the author for a nominal charge.

Reference Notes

1. Olkin, I., & Siotani, M. *Asymptotic distribution of functions of a correlation matrix* (Tech. Rep. 6). Stanford, Calif.: Stanford University, Laboratory for Quantitative Research in Education, 1964.
2. Steiger, J. H. *Multivariate analysis for correlations* (Tech. Rep. 78-35). Vancouver: University of British Columbia, Institute of Applied Mathematics and Statistics, September 1978.
3. Steiger, J. H. *Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results.* Paper presented at the meeting of the Psychometric Society, Monterey, Calif., June 1979.

References

Browne, M. W. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 1974, *8*, 1–24.

Browne, M. W. The analysis of patterned correlation matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology*, 1977, *30*, 113–124.

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

Ferguson, G. A. *Statistical analysis in psychology & education* (4th ed.). New York: McGraw-Hill, 1976.

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

Glass, G. V, & Stanley, J. C. *Statistical methods in education and psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1970.

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

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, 1973, *82*, 345–362.

Kenny, D. A. Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, 1975, *82*, 887–903.

Larzelere, R. E., & Mulaik, S. A. Single-sample tests for many correlations. *Psychological Bulletin*, 1977, *84*, 557–569.

McDonald, R. P. Testing pattern hypotheses for covariance matrices. *Psychometrika*, 1974, *39*, 189–201.

McDonald, R. P. Testing pattern hypotheses for correlation matrices. *Psychometrika*, 1975, *40*, 253–255.

McNemar, Q. *Psychological statistics* (4th ed.). New York: Wiley, 1969.

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

Olkin, I. Correlations revisited. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis.* Chicago: Rand McNally, 1967.

Pearson, K., & Filon, L. N. G. Mathematical contributions to the theory of evolution: IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London, Series A*, 1898, *191*, 229–311.

Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and the mathematical bases.* New York: McGraw-Hill, 1940.

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

Received October 30, 1978 ■