Far East Journal of Theoretical Statistics

http://www.pphmj.com
http://dx.doi.org/10.17654/TS053060333
Volume 53, Number 6, 2017, Pages 333-348 **ISSN: 0972-0863**

CONFIDENCE INTERVAL ESTIMATION FOR A DIFFERENCE BETWEEN TWO INTRACLASS CORRELATION COEFFICIENTS UNDER UNEQUAL FAMILY SIZES

Madhusudan Bhandary and Koji Fujiwara

Department of Mathematics
Columbus State University
4225 University Avenue
Columbus, GA 31907, U. S. A.

Department of Statistics
North Dakota State University
P. O. Box 5575, Waldron 201
Fargo, ND 58105, U. S. A.

Abstract

Confidence intervals (based on the F-distribution and the standard normal distribution) for the difference between two intraclass correlation coefficients under unequal family sizes based on two independent multinormal samples have been proposed. It has been found that the confidence interval based on the F-distribution produces better results than the confidence interval based on the standard normal distribution in terms of shorter average length. The coverage probability of the interval based on the F-distribution is competitive with that based on the standard normal distribution. An example with data is presented.

Received: August 23, 2017; Accepted: October 23, 2017

2010 Mathematics Subject Classification: 62H12.

Keywords and phrases: Z-test, F-distribution, intraclass correlation coefficient, confidence interval.

1. Introduction

Consider estimation of the intraclass correlation coefficient for blood pressure of p children in each of n families. The p measurements on a family provide $p(p-1)/2$ unique pairs of observations. From the n families, we generate a total of $np(p-1)/2$ unique pairs from which a correlation coefficient can be computed in the usual way. The correlation coefficient thus computed is called *intraclass correlation coefficient*.

Several authors have studied statistical inference concerning intraclass correlation coefficient based on a single multinormal sample: Scheffe [12], Rao [8], Rosner et al. [9, 10], Donner and Bull [2], Srivastava [13], Srivastava and Keen [15], Konishi [6], Gokhale and Sengupta [4] and Sengupta [11].

Donner and Bull [2] discussed the likelihood ratio test for testing the equality of two intraclass correlation coefficients based on two independent multinormal samples under equal family sizes. Konishi and Gupta [7] proposed a modified likelihood ratio test and derived its asymptotic null distribution. They also discussed another test procedure based on a modification of Fisher's Z-transformation following Konishi [6].

Huang and Sinha [5] considered an optimum invariant test for the equality of intraclass correlation coefficients under equal family sizes for more than two intraclass correlation coefficients based on independent samples from several multinormal distributions.

For unequal family sizes, Young and Bhandary [17] proposed the likelihood ratio test, large sample Z-test and large sample Z^*-test for the equality of two intraclass correlation coefficients based on two independent multinormal samples.

For several populations and unequal family sizes, Bhandary and Alam [1] proposed the likelihood ratio test and large sample ANOVA test for the equality of several intraclass correlation coefficients based on several independent multinormal samples. Donner and Zou [3] proposed asymptotic test for the equality of dependent intraclass correlation coefficients under unequal family sizes.

But none of the above authors derived any confidence interval estimator for the difference between two intraclass correlation coefficients under unequal family sizes. In this paper, it is considered confidence interval estimators for the difference between two intraclass correlation coefficients based on two independent multinormal samples under unequal family sizes.

It has been carried out conditional analysis assuming family sizes fixed though unequal.

It could be of interest to estimate the difference in heritability in blood pressure or cholesterol or lung etc., between families in Native American/or/ White races and the families in Asian races and therefore we need to develop an interval estimator for the difference between two intraclass correlation coefficients under unequal family sizes.

In Section 2, confidence interval estimators for the difference between two intraclass correlation coefficients under unequal family sizes have been proposed. These interval estimators are compared in Section 3 using a simulation technique. It is found on the basis of simulation study that the confidence interval estimator based on the *F*-distribution produces better result than the confidence interval estimator based on the standard normal distribution in terms of shorter average length. The coverage probability of the interval based on the *F*-distribution is competitive with that based on the standard normal distribution. An example with data is presented in Section 4.

2. Proposed Confidence Intervals

2.1. Interval based on standard normal distribution

Let $X_i = (x_{i1}, x_{i2}, ..., x_{ip_i})'$ be a $p_i \times 1$ vector of observations from *i*th family; $i = 1, 2, ..., k_1$. The mean vector and the covariance matrix for the familial data are:

$$\underset{\sim}{\mu}_i = \mu_1 \underset{\sim}{1}_i \text{ and } \underset{p_i \times p_i}{\Sigma_{1i}} = \sigma_1^2 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_1 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_1 & \rho_1 & \cdots & 1 \end{pmatrix}, \tag{2.1}$$

where $\underset{\sim}{1}_i$ is a $p_i \times 1$ vector of 1's, μ_1 $(-\infty < \mu_1 < \infty)$ is the common mean and σ_1^2 $(\sigma_1^2 > 0)$ is the common variance of members of the family and ρ_1 is the intraclass correlation coefficient with $\max_{1 \le i \le k}\left(-\frac{1}{p_i - 1}\right) \le \rho_1 \le 1.$

It is assumed that $\underset{\sim}{x}_i \sim N_{p_i}\left(\underset{\sim}{\mu}_i, \Sigma_{1i}\right)$; $i = 1, ..., k_1$, where N_{p_i} represents p_i-variate normal distribution and $\underset{\sim}{\mu}_i, \Sigma_{1i}$'s are defined in (2.1). Let

$$\underset{\sim}{u}_i = (u_{i1}, u_{i2}, ..., u_{ip_i})' = Q \underset{\sim}{x}_i, \tag{2.2}$$

where Q is Helmert's orthogonal matrix.

Under this transformation, $\underset{\sim}{u}_i \sim N_{p_i}(\underset{\sim}{\mu}_i^*, \Sigma_{1i}^*)$; $i = 1, ..., k_1$, where

$$\underset{p_i \times 1}{\underset{\sim}{\mu}_i^*} = (\mu_1, 0, ..., 0)' \text{ and } \Sigma_{1i}^* = \sigma_1^2 \begin{pmatrix} \eta_i & 0 & \cdots & 0 \\ 0 & 1 - \rho_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 - \rho_1 \end{pmatrix}$$

and $\eta_i = p_i^{-1}\{1 + (p_i - 1)\rho_1\}$.

Srivastava [13] gives estimator of ρ_1 and σ_1^2 under unequal family sizes which are good substitute for the maximum likelihood estimator and are given by the following:

$$\hat{\rho}_1 = 1 - \frac{\hat{\gamma}^2}{\hat{\sigma}_1^2},$$

$$\hat{\sigma}_1^2 = (k_1 - 1)^{-1} \sum_{i=1}^{k_1} (u_{i1} - \hat{\mu}_1)^2 + k_1^{-1} \hat{\gamma}^2 \left(\sum_{i=1}^{k_1} a_i \right), \tag{2.3}$$

where

$$\hat{\gamma}^2 = \frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2}{\sum_{i=1}^{k_1} (p_i - 1)}, \quad \hat{\mu}_1 = k_1^{-1} \sum_{i=1}^{k_1} u_{i1} \text{ and } a_i = 1 - p_i^{-1}.$$

Srivastava and Katapa [14] derived the asymptotic distribution of $\hat{\rho}_1$. They have shown that $\hat{\rho}_1 \sim N(\rho_1, Var/k_1)$ asymptotically where

$$Var = 2(1 - \rho_1)^2 \left\{ (\overline{p} - 1)^{-1} + c^2 - 2(1 - \rho_1)(\overline{p} - 1)^{-1} k_1^{-1} \sum_{i=1}^{k_1} a_i \right\},$$

$k_1 =$ number of families in the sample,

$$\overline{p} = k_1^{-1} \sum_{i=1}^{k_1} p_i,$$

$$c^2 = 1 - 2(1 - \rho_1)^2 k_1^{-1} \sum_{i=1}^{k_1} a_i$$

$$+ (1 - \rho_1)^2 \left[k_1^{-1} \sum_{i=1}^{k_1} a_i + (\overline{p} - 1)^{-1} \left(k_1^{-1} \sum_{i=1}^{k_1} a_i \right)^2 \right]. \tag{2.4}$$

Consider the two sample problem with k_1 and k_2 families from each population.

Let $\underset{\sim}{y}_j = (y_{j1}, y_{j2}, \ldots, y_{jq_j})'$ be a $q_j \times 1$ vector of observations from jth family in the second population; $j = 1, \ldots, k_2$ and

$$\underset{\sim}{y}_j \sim N_{q_j}(\underset{\sim}{\mu}_{2j}, \Sigma_{2j}), \tag{2.5}$$

where

$$\underset{\sim}{\mu}_{2j} = \mu_2 \underset{\sim}{1}_j, \quad \Sigma_{2j} = \sigma_2^2 \begin{pmatrix} 1 & \rho_2 & \cdots & \rho_2 \\ \rho_2 & 1 & \cdots & \rho_2 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix} \tag{2.6}$$

and $\max\limits_{1 \le j \le k_2} \left(-\dfrac{1}{q_j - 1} \right) \le \rho_2 \le 1.$

Using Helmert's transformation, we can transform the data vector from $\underset{\sim}{x}_i$ to $\underset{\sim}{u}_i$ and $\underset{\sim}{y}_j$ to $\underset{\sim}{v}_j$ as follows:

$$\underset{\sim}{u}_i = (u_{i1}, u_{i2}, \ldots, u_{ip_i})' \sim N_{p_i}(\underset{\sim}{\mu}_{1i}^*, \Sigma_{1i}^*); \quad i = 1, \ldots, k_1$$

and

$$\underset{\sim}{v}_j = (v_{j1}, v_{j2}, \ldots, v_{jq_j})' \sim N_{q_j}(\underset{\sim}{\mu}_{2j}^*, \Sigma_{2j}^*); \quad j = 1, \ldots, k_2, \tag{2.7}$$

where

$$\underset{\sim}{\mu}_{1i}^* = (\mu_1, 0, \ldots, 0)', \quad \Sigma_{1i}^* = \sigma_1^2 \begin{pmatrix} \eta_i & 0 & \cdots & 0 \\ 0 & 1-\rho_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1-\rho_1 \end{pmatrix},$$

$$\eta_i = p_i^{-1}\{1 + (p_i - 1)\rho_1\},$$

$$\underset{\sim}{\mu}^*_{2j} = (\mu_2, 0, ..., 0)', \quad \Sigma^*_{2j} = \sigma_2^2 \begin{pmatrix} \xi_j & 0 & \cdots & 0 \\ 0 & 1-\rho_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1-\rho_2 \end{pmatrix}$$

and $\xi_j = q_j^{-1}\{1 + (q_j - 1)\rho_2\}$.

Assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Srivastava and Katapa [14] showed that

$$Z^* = \frac{\hat{\rho}_1 - \hat{\rho}_2 - (\rho_1 - \rho_2)}{\sqrt{\frac{S_1^2}{k_1} + \frac{S_2^2}{k_2}}} \sim N(0, 1) \text{ asymptotically,} \tag{2.8}$$

where

$$S_1^2 = 2(1-\hat{\rho}_1)^2\left\{(\overline{p}-1)^{-1} + c^2 - 2(1-\hat{\rho}_1)(\overline{p}-1)^{-1}k_1^{-1}\sum_{i=1}^{k_1} a_i\right\},$$

$$\overline{p} = k_1^{-1}\sum_{i=1}^{k_1} p_i,$$

$$c^2 = 1 - 2(1-\hat{\rho}_1)^2 k_1^{-1}\sum_{i=1}^{k_1} a_i$$

$$+ (1-\hat{\rho}_1)^2\left[k_1^{-1}\sum_{i=1}^{k_1} a_i + (\overline{p}-1)^{-1}\left(k_1^{-1}\sum_{i=1}^{k_1} a_i\right)^2\right]$$

and $\hat{\rho}_1$ is given by (2.3).

Similarly, S_2^2 is obtained with $\hat{\rho}_1$ replaced by $\hat{\rho}_2$ and k_1 replaced by k_2.

The square of the denominator of (2.8) is a consistent estimator of $Var(\hat{\rho}_1 - \hat{\rho}_2)$. Using the expression (2.8), a $100(1-\alpha)\%$ confidence interval for $\rho_1 - \rho_2$ is

$$(\hat{\rho}_1 - \hat{\rho}_2) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{k_1} + \frac{S_2^2}{k_2}}. \tag{2.9}$$

Note that

$$Z = \frac{\hat{\rho}_1 - \hat{\rho}_2 - (\rho_1 - \rho_2)}{S\sqrt{\frac{1}{k_1} + \frac{1}{k_2}}} \sim N(0, 1),$$

where the pooled variance estimator $S^2 = \frac{(k_1 - 1)S_1^2 + (k_2 - 1)S_2^2}{k_1 + k_2 - 2}$.

When $k_1 = k_2$, $\sqrt{\frac{S_1^2}{k_1} + \frac{S_2^2}{k_2}} = S\sqrt{\frac{1}{k_1} + \frac{1}{k_2}}$ and hence the confidence intervals produced by Z^* and Z are same. In the simulation section, data has been generated for $k_1 = k_2$ and that is why only Z^* has been used for confidence interval.

2.2. Interval based on F-distribution

Let

$$n_1 = \sum_{i=1}^{k_1}(p_i - 1) \text{ and } n_2 = \sum_{j=1}^{k_2}(q_j - 1),$$

$$\hat{\gamma}_1^2 = \frac{\sum_{i=1}^{k_1}\sum_{r=2}^{p_i} u_{ir}^2}{n_1} \text{ and } \hat{\gamma}_2^2 = \frac{\sum_{j=1}^{k_2}\sum_{s=2}^{q_j} v_{js}^2}{n_2}.$$

Note that

$$\frac{n_1\hat{\gamma}_1^2}{\sigma^2(1-\rho_1)} \sim \chi^2_{n_1} \tag{2.10}$$

and

$$\frac{n_2\hat{\gamma}_2^2}{\sigma^2(1-\rho_2)} \sim \chi^2_{n_2}, \tag{2.11}$$

where χ^2_n denotes the chi-square distribution with n degrees of freedom.

The exact distribution of the statistic

$$F = \left(\frac{\hat{\gamma}_1^2}{\hat{\gamma}_2^2}\right) \cdot \left(\frac{1-\rho_2}{1-\rho_1}\right) \text{ is } F_{n_1, n_2} \tag{2.12}$$

with degrees of freedom n_1 and n_2, respectively.

The $100\left(1-\frac{\alpha}{2}\right)\%$ confidence interval for $\frac{1-\rho_2}{1-\rho_1}$ is:

$$(F_{1-\frac{\alpha}{4}; n_1, n_2})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right) < \frac{1-\rho_2}{1-\rho_1} < (F_{\frac{\alpha}{4}; n_1, n_2})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right). \tag{2.13}$$

An approximate $100\left(1-\frac{\alpha}{2}\right)\%$ confidence interval for $\rho_1 - \rho_2$ is:

$$\hat{\rho}^A < \rho_1 - \rho_2 < \hat{\rho}^B,$$

$$\hat{\rho}^A = \left((F_{1-\frac{\alpha}{4}; n_1, n_2})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right) - 1\right)(1-\hat{\rho}_1),$$

$$\hat{\rho}^B = \left((F_{\frac{\alpha}{4}; n_1, n_2})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right) - 1\right)(1-\hat{\rho}_1). \tag{2.14}$$

Again, from (2.13), one can also write

$$(F_{1-\frac{\alpha}{4}; n_2, n_1})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right) < \frac{1-\rho_1}{1-\rho_2} < (F_{\frac{\alpha}{4}; n_2, n_1})\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right). \tag{2.15}$$

A second approximate $100\left(1-\frac{\alpha}{2}\right)\%$ confidence interval for $\rho_1 - \rho_2$ is:

$$\hat{\rho}^C < \rho_1 - \rho_2 < \hat{\rho}^D,$$

$$\hat{\rho}^C = \left(1 - \left(F_{\frac{\alpha}{4}; n_2, n_1}\right)\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right)\right)(1 - \hat{\rho}_2),$$

$$\hat{\rho}^D = \left(1 - \left(F_{1-\frac{\alpha}{4}; n_2, n_1}\right)\left(\frac{\hat{\gamma}_2^2}{\hat{\gamma}_1^2}\right)\right)(1 - \hat{\rho}_2). \tag{2.16}$$

Therefore, an approximate $100(1-\alpha)\%$ confidence interval for $\rho_1 - \rho_2$ is as follows:

$$(\hat{\rho}^A, \hat{\rho}^B) \cap (\hat{\rho}^C, \hat{\rho}^D). \tag{2.17}$$

Because

$$\begin{aligned} &P((\hat{\rho}^A, \hat{\rho}^B) \cap (\hat{\rho}^C, \hat{\rho}^D)) \\ &= P((\hat{\rho}^A, \hat{\rho}^B)) + P((\hat{\rho}^C, \hat{\rho}^D)) - P((\hat{\rho}^A, \hat{\rho}^B) \cup (\hat{\rho}^C, \hat{\rho}^D)) \\ &\geq 1 - \frac{\alpha}{2} + 1 - \frac{\alpha}{2} - 1 \text{ (by (2.14) and (2.16))} \\ &= 1 - \alpha. \end{aligned}$$

This approximate interval is conservative. Performance of the intervals given by (2.9) and (2.17) in terms of average lengths and coverage probabilities will be discussed in the next section using simulated data.

3. Simulation Results

Multivariate normal random vectors were generated using R. Two cases of five $(k_1 = k_2 = 5)$ and thirty $(k_1 = k_2 = 30)$ vectors of family data were created for each of the two populations. The family size distribution was truncated to maintain the family size at a minimum of 2 siblings and a maximum of 15 siblings. The previous research in simulating family sizes (Rosner et al. [9] and Srivastava and Keen [15]) used a negative binomial distribution with a mean = 2.86 and a probability of success of 0.483. Here, it was set at a mean = 2.86 and a theta = 41.2552 which were the same as those of the previous researchers. The parameters ρ_1 and ρ_2 took combinations over the range of values from 0.1 to 0.9. There were 10,000 Monte Carlo samples for each pair of ρ_1 and ρ_2. The average length and coverage probability of each interval at $\alpha = 0.05$ were noted (Table 1). On the basis of our study, the interval based on the *F*-distribution showed shorter average length than the interval based on the standard normal distribution. The coverage probability of the interval based on the *F*-distribution is competitive with the coverage probability of the interval based on the standard normal distribution (Table 2).

Table 1. Comparison of length $(\alpha = 0.05)$

		$k = 5$		$k = 30$	
ρ_1	ρ_2	Z	F_0	Z	F_0
0.1	0.4	1.3791	1.0303	0.5810	0.5565
0.1	0.5	1.3452	0.9597	0.5624	0.5175
0.2	0.3	1.3954	1.0399	0.5968	0.5630
0.2	0.7	1.2439	0.7782	0.5095	0.4090
0.3	0.3	1.3890	1.0018	0.5969	0.5271
0.3	0.5	1.3378	0.8874	0.5707	0.4553
0.3	0.7	1.2347	0.7479	0.5097	0.3769
0.3	0.9	1.0617	0.5574	0.4381	0.2967
0.4	0.2	1.3763	0.9925	0.5881	0.5262
0.4	0.4	1.3487	0.8975	0.5791	0.4567
0.4	0.5	1.3177	0.8439	0.5604	0.4202
0.4	0.7	1.2181	0.7106	0.4989	0.3423
0.4	0.9	1.0351	0.5201	0.4256	0.2626
0.5	0.1	1.3484	0.9653	0.5631	0.5179
0.5	0.4	1.3188	0.8439	0.5606	0.4193

0.5	0.5	1.2863	0.7954	0.5415	0.3835
0.5	0.7	1.1790	0.6607	0.4765	0.3073
0.5	0.9	0.9977	0.4782	0.4002	0.2288
0.6	0.3	1.2944	0.8254	0.5426	0.4163
0.6	0.6	1.1929	0.6711	0.4836	0.3102
0.6	0.9	0.9347	0.4270	0.3607	0.1933
0.7	0.3	1.2349	0.7469	0.5094	0.3758
0.7	0.7	1.0527	0.5379	0.4021	0.2348
0.7	0.9	0.8458	0.3699	0.3067	0.1562
0.8	0.5	1.0957	0.5759	0.4370	0.2663
0.8	0.8	0.8590	0.3919	0.2973	0.1587
0.8	0.9	0.7210	0.3020	0.2398	0.1188
0.9	0.9	0.5581	0.2206	0.1637	0.0800

Table 2. Comparison of coverage probability $(\alpha = 0.05)$

		$k = 5$		$k = 30$	
ρ_1	ρ_2	Z	F_0	Z	F_0
0.1	0.1	0.9808	0.9751	0.9782	0.9724
0.1	0.2	0.9741	0.9741	0.9663	0.9748
0.1	0.3	0.9697	0.9636	0.9577	0.9710
0.1	0.4	0.9620	0.9459	0.9568	0.9611
0.1	0.5	0.9573	0.9081	0.9632	0.9446
0.2	0.1	0.9762	0.9732	0.9616	0.9726
0.2	0.2	0.9683	0.9754	0.9489	0.9757
0.2	0.3	0.9607	0.9713	0.9488	0.9726
0.2	0.4	0.9545	0.9530	0.9427	0.9655
0.2	0.5	0.9498	0.9185	0.9453	0.9515
0.3	0.1	0.9676	0.9627	0.9577	0.9690
0.3	0.2	0.9617	0.9723	0.9455	0.9754
0.3	0.3	0.9528	0.9746	0.9432	0.9737
0.3	0.4	0.9380	0.9672	0.9445	0.9717
0.3	0.5	0.9331	0.9395	0.9424	0.9603

0.3	0.6	0.9320	0.8832	0.9386	0.8717
0.4	0.1	0.9653	0.9434	0.9567	0.9630
0.4	0.2	0.9578	0.9622	0.9459	0.9658
0.4	0.3	0.9421	0.9681	0.9392	0.9716
0.4	0.4	0.9340	0.9732	0.9391	0.9756
0.4	0.5	0.9233	0.9621	0.9360	0.9691
0.4	0.6	0.9266	0.9142	0.9404	0.9494
0.4	0.7	0.9222	0.8394	0.9389	0.8856
0.5	0.1	0.9570	0.9136	0.9641	0.9492
0.5	0.2	0.9491	0.9255	0.9431	0.9508
0.5	0.3	0.9341	0.9394	0.9382	0.9598
0.5	0.4	0.9265	0.9647	0.9396	0.9714
0.5	0.5	0.9180	0.9770	0.9392	0.9756
0.5	0.6	0.9071	0.9517	0.9416	0.9651
0.5	0.7	0.9046	0.8757	0.9420	0.9221
0.6	0.3	0.9387	0.8922	0.9399	0.9336
0.6	0.4	0.9220	0.9221	0.9378	0.9505
0.6	0.5	0.9117	0.9537	0.9416	0.9635
0.6	0.6	0.9106	0.9766	0.9455	0.9769
0.6	0.7	0.9129	0.9299	0.9478	0.9560
0.7	0.5	0.9127	0.8843	0.9419	0.9214
0.7	0.6	0.9107	0.9318	0.9458	0.9598
0.7	0.7	0.9158	0.9751	0.9514	0.9717
0.7	0.8	0.9201	0.8883	0.9495	0.9311
0.8	0.5	0.9064	0.7655	0.9435	0.8141
0.8	0.6	0.9096	0.8112	0.9450	0.8595
0.8	0.7	0.9185	0.8892	0.9528	0.9284
0.8	0.8	0.9383	0.9748	0.9532	0.9729
0.8	0.9	0.9341	0.7805	0.9523	0.8374
0.9	0.8	0.9384	0.7863	0.9514	0.8378
0.9	0.9	0.9677	0.9771	0.9620	0.9752

4. Example with Real Life Data

In this section, a comparison of the two intervals is going to be done using data from Srivastava and Katapa [14] (Table 3).

Table 3. Values of pattern intensity on soles of feet in 14 families in two groups

Sample	Family#	Mother	Father	#Siblings	Data
A	1	2	3	2	2, 2
A	2	2	3	2	2, 3
A	6	4	3	3	4, 3, 3
A	8	3	7	7	2, 4, 7, 4, 4, 7, 8
A	11	5	6	4	5, 3, 4, 4
A	12	2	4	2	2, 4
A	14	2	3	3	2, 2, 2
Sample	Family#	Mother	Father	#Siblings	Data
B	3	2	3	3	2, 2, 2
B	4	2	4	5	2, 2, 2, 2, 2
B	5	6	7	2	6, 6
B	7	4	3	7	2, 2, 3, 6, 3, 5, 4
B	9	5	5	2	5, 6
B	10	5	4	3	4, 5, 4
B	13	6	3	4	4, 3, 3, 3

Here $k_1 = 7$ and $k_2 = 7$.

The approximate 95% confidence intervals for the difference between two intraclass correlation coefficients are (–0.70297, 0.0367) using the F-distribution approximation and (–1.2981, 0.23691) using the standard normal approximation.

The length of the 95% confidence interval based on the F-distribution approximation is much shorter than the length of the 95% confidence interval based on the standard normal distribution.

Acknowledgement

The authors thank the anonymous referee for very careful reading of the paper and for making some valuable suggestions that helped to improve its original version.

References

[1] M. Bhandary and M. K. Alam, Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations, Comm. Statist. Theory Methods 29(4) (2000), 755-768.

[2] A. Donner and S. Bull, Inferences concerning a common intraclass correlation coefficient, Biometrics 39 (1983), 771-775.

[3] A. Donner and G. Zou, Testing the equality of dependent intraclass correlation coefficients, The Statistician 51(3) (2002), 367-379.

[4] D. V. Gokhale and A. Sengupta, Optimal tests for the correlation coefficient in a symmetric multivariate normal population, J. Statist. Plann. Inference 14 (1986), 263-268.

[5] W. Huang and B. K. Sinha, On optimum invariant tests of equality of intraclass correlation coefficients, Ann. Inst. Statist. Math. 45(3) (1993), 579-597.

[6] S. Konishi, Normalizing and variance stabilizing transformations for intraclass correlations, Ann. Inst. Statist. Math. 37 (1985), 87-94.

[7] S. Konishi and A. K. Gupta, Testing the equality of several intraclass correlation coefficients, J. Statist. Plann. Inference 21 (1989), 93-105.

[8] C. R. Rao, Linear Statistical Inference and its Applications, Wiley, New York, 1973.

[9] B. Rosner, A. Donner and C. H. Hennekens, Estimation of intraclass correlation from familial data, Appl. Stat. 26 (1977), 179-187.

[10] B. Rosner, A. Donner and C. H. Hennekens, Significance testing of interclass correlations from familial data, Biometrics 35 (1979), 461-471.

[11] A. Sengupta, On loss of power under additional information - an example, Scand. J. Statist. 15 (1988), 25-31.

[12] H. Scheffe, The Analysis of Variance, Wiley, New York, 1959.

[13] M. S. Srivastava, Estimation of interclass correlations in familial data, Biometrika 71 (1984), 177-185.

[14] M. S. Srivastava and R. S. Katapa, Estimation of intraclass correlations in familial data, Canadian Journal of Statistics 14 (1986), 29-42.

[15] M. S. Srivastava and K. J. Keen, Estimation of the interclass correlation coefficient, Biometrika 75 (1988), 731-739.

[16] R. Velu and M. B. Rao, Estimation of parent-offspring correlation, Biometrika 77(3) (1990), 557-562.

[17] D. Young and M. Bhandary, Test for equality of intraclass correlation coefficients under unequal family sizes, Biometrics 54(4) (1998), 1363-1373.