Thailand Statistician
July 2015; 13(2): 185-190
http://statassoc.or.th
Contributed paper

On Testing Equality of Two Correlation Coefficients

Syed Ejaz Ahmed [a] Arjun Kumar Gupta [b] and Ngoc Nguyen* [c]

[a] Department of Mathematics and Statistics, Brock University St Catharines, ON L2S 3A1, Canada.

[b] Department of Mathematics and Statistics, Bowling Green State University Bowling Green, Ohio 43403, U.S.A.

[c] Department of Mathematics, Western Kentucky University Bowling Green, Kentucky 42101, U.S.A.

*Corresponding author; e-mail: ngoc.nguyen@wku.edu

Received: 2 October 2014
Accepted: 10 February 2015

Abstract

A simple significance test for comparing two correlation coefficients estimated from independent samples is given. The test proposed is exact under very mild conditions on covariance matrices. Confidence interval of the difference in correlation coefficients is constructed for special cases.

Keywords: correlation coefficients, bivariate normal, hypothesis testing, confidence intervals.

1. Introduction

Quite often the researchers have two independent samples, each randomly selected from a bivariate normal population. If the research workers can assume that the two correlations are equal, then it is desirable to obtain a pooled estimate of the common population correlation coefficient, and the pooled estimator is more efficient than any of the individual estimates. However, the test used is usually an asymptotic one, based on the Fisher z-transformation, for example see Kendall and Stuart [1]. Here we present an exact test under mild conditions. In fact, in a regressive or conditional set-up, the proposed condition is quite natural if x_1 and x_2 are controlled factors and $y = y_1 = y_2$ (see Section 2).

Let $(X_{ji}, Y_{ji}), i = 1, \dots, n$ and $j = 1, 2$ be two independent samples of size n from the bivariate normal populations,

$$N_2\left\{\begin{pmatrix}\mu_{x_1}\\ \mu_{y_1}\end{pmatrix}, \begin{pmatrix}\sigma_{x_1}^2 & \rho_1\sigma_{x_1}\sigma_{y_1}\\ \rho_1\sigma_{x_1}\sigma_{y_1} & \sigma_{y_1}^2\end{pmatrix}\right\}$$

and

$$N_2\left\{\begin{pmatrix}\mu_{x_2}\\ \mu_{y_2}\end{pmatrix}, \begin{pmatrix}\sigma_{x_2}^2 & \rho_2\sigma_{x_2}\sigma_{y_2}\\ \rho_2\sigma_{x_2}\sigma_{y_2} & \sigma_{y_2}^2\end{pmatrix}\right\},$$

respectively.

The hypothesis of interest is

$$H_0 : \rho_1 = \rho_2 \qquad (1)$$

where $corr(X_1, Y_1) = \rho_1$ and $corr(X_2, Y_2) = \rho_2$. Several authors have studied tests for the hypothesis H_0 in relation (1) including Hotelling [2] and Aitken et al. [3]. In the next section we derive an exact test for H_0. In Section 3, we give results for confidence interval of the difference between two correlation coefficients.

2. Paired t-test for H_0

Let

$$W = X_1 - X_2,$$

and

$$Z = Y_1 + Y_2, \qquad (2)$$

then

$$(W,Z)\sim N_2\left\{\begin{pmatrix}\mu_{x_1}-\mu_{x_2}\\ \mu_{y_1}+\mu_{y_2}\end{pmatrix},\begin{pmatrix}\sigma_{x_1}^2+\sigma_{x_2}^2 & \rho(w,z)\sigma_w\sigma_z\\ \rho(w,z)\sigma_w\sigma_z & \sigma_{y_1}^2+\sigma_{y_2}^2\end{pmatrix}\right\}, \tag{3}$$

where

$$\sigma_w^2=\sigma_{x_1}^2+\sigma_{x_2}^2 \text{ and } \sigma_z^2=\sigma_{y_1}^2+\sigma_{y_2}^2,$$

and

$$\rho(w,z)=\frac{\sigma_{x_1}\sigma_{y_1}}{\sigma_w\sigma_z}\rho_1-\frac{\sigma_{x_2}\sigma_{y_2}}{\sigma_w\sigma_z}\rho_2.$$

Consequently if $\sigma_{x_1}\sigma_{y_1}=\sigma_{x_2}\sigma_{y_2}$ then

$$H_0:\rho_1=\rho_2 \Leftrightarrow H_0':\rho(w,z)=0.$$

Kshirsagar and Radhakrishnan [4] have given a likelihood ratio test fortesting $\sigma_{x_1}\sigma_{y_1}=\sigma_{x_2}\sigma_{y_2}$ when H_0 is true. Now estimate $\rho(w,z)$ by

$$r(w,z)=\frac{\sum_{i=1}^n(w_i-\bar{w})(z_i-\bar{z})}{\sqrt{\sum_{i=1}^n(w_i-\bar{w})^2\sum_{i=1}^n(z_i-\bar{z})^2}}. \tag{4}$$

For testing H_0: $\rho(w,z)=0$ vs H_1: $\rho(w,z)\neq 0$, reject the null hypothesis H_0 (see e.g. Pitman [5]) if

$$\frac{\sqrt{n-2}|r(w,z)|}{\sqrt{1-r^2(w,z)}}>t_{(n-2)}(\alpha), \tag{5}$$

and for testing H_0: $\rho(w,z)=0$ vs H_1: $\rho(w,z)>0$, reject the null hypothesis H_0 if

$$\frac{\sqrt{n-2}r(w,z)}{\sqrt{1-r^2(w,z)}}>t_{(n-2)}(2\alpha). \tag{6}$$

Furthermore, the corresponding asymptotic normality of $z_j = \frac{1}{2} log \frac{1+r(x_j,y_j)}{1-r(x_j,y_j)}, j = 1, 2$ also holds and approximate test based on $(z_1 - z_2)/\sqrt{\frac{2}{n-3}}$ can be carried out with large samples.

3. Confidence intervals for $\rho_1 - \rho_2$

As presented in Section 2, $\rho(w, z)$ can be used to test H0: $\rho_1 = \rho_2$. However, in general, it cannot be used to obtain confidence intervals or bounds for the difference $\rho_1 - \rho_2$ since it will involve unknown parameters $\mu_{x_j}, \mu_{y_j}, \sigma_{x_j}$ and $\sigma_{y_j}, j = 1, 2$. In this section, we present the confidence intervals for $\rho_1 - \rho_2$ in the following case

$$(X_1, Y_1) \sim N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \right\},$$

$$(X_2, Y_2) \sim N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix} \right\}. \tag{7}$$

Hence,

$$(W, Z) \sim N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & \rho_1 - \rho_2 \\ \rho_1 - \rho_2 & 2 \end{pmatrix} \right\}. \tag{8}$$

By Anderson [6], $n(\bar{X} - \mu)'^{\Sigma^{-1}}(\bar{X} - \mu) \sim \chi^2_{(2)}$ where a sample of size n is chosen from $N_2(\mu, \Sigma)$. Then a (1-2α)100% confidence interval for $\rho_1 - \rho_2$ is readily derived

$$\chi^2_{2,1-\alpha} < n(\bar{w} \quad \bar{z})' \begin{pmatrix} 2 & \rho_1 - \rho_2 \\ \rho_1 - \rho_2 & 2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{w} \\ \bar{z} \end{pmatrix} < \chi^2_{2,\alpha}$$

$$\Leftrightarrow \chi^2_{2,1-\alpha} < \frac{n}{4-(\rho_1-\rho_2)^2} [2(\bar{w}^2 + \bar{z}^2) + 2\bar{w}\bar{z}(\rho_1 - \rho_2)] < \chi^2_{2,\alpha}. \tag{9}$$

The relationship (9) is equivalent to

$$\begin{cases} L(\rho_1 - \rho_2)^2 - 2n\bar{w}\bar{z}(\rho_1 - \rho_2) + 2n(\bar{w}^2 + \bar{z}^2) - 4L > 0 \\ R(\rho_1 - \rho_2)^2 - 2n\bar{w}\bar{z}(\rho_1 - \rho_2) + 2n(\bar{w}^2 + \bar{z}^2) - 4R < 0 \end{cases} \tag{10}$$

where $L = \chi^2_{2,1-\alpha}$ and $R = \chi^2_{2,\alpha}$. Range of $\rho_1 - \rho_2$ that satisfies (10) constitutes a corresponding (1-2α)100% confidence interval.

The result can be readily generalized to the case where the mean vectors for the two bivariate distributions are known and the variances are known and satisfy the condition $\sigma_{x_1}\sigma_{y_1} = \sigma_{x_2}\sigma_{y_2}$.

Alternatively, we can use Fisher z-transformation:

$$z_j = \frac{1}{2}log\frac{1+r(x_j,y_j)}{1-r(x_j,y_j)}, j = 1,2.$$

If n is sufficiently large, $z_j \to N(F(\rho_j), \frac{1}{\sqrt{n-3}})$

where $F(\rho_j) = \frac{1}{2}log\frac{1+\rho(x_j,y_j)}{1-\rho(x_j,y_j)}, j = 1,2,$ and thus $z_1 - z_2 \to N(F(\rho_1) - F(\rho_2), \sqrt{\frac{2}{n-3}})$. From this, we can construct confidence intervals for $F(\rho_1) - F(\rho_2)$.

For illustration purpose, we simulate random samples of different sizes from the bivariate distributions given in (7) with different values of ρ_1 and ρ_2 and construct 80% and 90% confidence intervals for $\rho_1 - \rho_2$ and $F(\rho_1) - F(\rho_2)$. The results are presented in Table 1. Note that the fourth column shows the exact confidence interval for $\rho_1 - \rho_2$ while the fifth column shows the asymptotic confidence interval for $F(\rho_1) - F(\rho_2)$.

Table 1. Confidence intervals.

			$\rho_1 - \rho_2$	$F(\rho_1) - F(\rho_2)$
$\rho_1 = 0.3$	n = 30	cl = 0.8	(-0.1063,1.989)	(-0.6572,0.0396)
$\rho_2 = 0.3$		cl = 0.9	(-0.5275,1.9997)	(-0.7565,0.1389)
	n = 40	cl = 0.8	(-0.0523,1.9954)	(-0.6202,-0.0250)
		cl = 0.9	(-0.4125,1.9977)	(-0.7051,0.0599)
$\rho_1 = 0.5$	n = 30	cl = 0.8	(0.0978,1.9932)	(-0.5289,0.1679)
$\rho_2 = 0.3$		cl = 0.9	(0.0124,1.9956)	(-0.6282,0.2672)
	n = 40	cl = 0.8	(0.1078,1.9924)	(-0.2481,0.3471)
		cl = 0.9	(0.0341,1.9930)	(-0.333,0.4310)
$\rho_1 = 0.8$	n = 30	cl = 0.8	(0.2786,1.9991)	(-0.1388,0.5580)
$\rho_1 = 0.3$		cl = 0.9	(-0.2489,1.9995)	(-0.2381,0.6573)
	n = 40	cl = 0.8	(0.3391,1.9978)	(0.4414,1.0366)
		cl = 0.9	(0.1352,1.9990)	(0.3565,1.1214)

All simulations were done using R statistical package. When $\rho_1 = \rho_2 = 0.3$, all confidence intervals contain zero, which indicates insignificant difference between two correlation coefficients. When $\rho_1 \neq \rho_2$, almost all confidence intervals do not contain zero, which indicates significant difference between two correlations. We can also see from the numerical results, the confidence intervals are quite wide with small sample sizes and large confidence levels.

4. Conclusion

In the paper, an exact test is derived under mild conditions to test equality of correlation coefficients of independent bivariate normal populations. The test pairs the values of the random variables assuming that two samples are of the same size. In many studies, however, such transformation may be natural. Confidence intervals for the difference of correlation coefficients are also presented for the special case.

Acknowledgments

The work of the first author was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 1, 4th edition, Macmillan: New York, 1977.

[2] Hotelling, H., The Selection of Variates for Use in Prediction with some Comments on the General Problem of Nuisance Parameters, *Ann.Math. Statist*., 1940; 11: 271-283.

[3] Aitken, M.A., Nelson, W.C. and Reinfurt, K.H., Tests for Correlation Matrices. *Biometrika*, 1968; 55: 327-334.

[4] Kshirsagar, A.M. and Radhakrishnan, R., A Likelihood Ratio Test for Interaction Among Variances, *Comm. Statist. - Theory Meth.*, 1989; 18: 4511-4517.

[5] Pitman, E.J. G., A Note on Normal Correlation, *Biometrika*, 1939; 31: 9-12.

[6] Anderson T.W., *An Introduction to Multivariate Analysis*, 2nd edition, Wiley: New York, 1984.