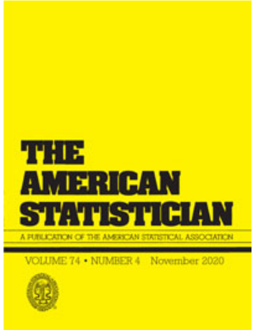

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Xinjie Hu, Aekyung Jung, and Gengsheng Qin (2020), “Interval Estimation for the Correlation Coefficient,” *The American Statistician*, 74:1, 29–36: Comment by Krishnamoorthy and Xia

Kalimuthu Krishnamoorthy & Yanping Xia

THE AMERICAN STATISTICIAN
2020, VOL. 74, NO. 4, 418: Letter to the Editor
https://doi.org/10.1080/00031305.2020.1829048

LETTER TO THE EDITOR

Xinjie Hu, Aekyung Jung, and Gengsheng Qin (2020), "Interval Estimation for the Correlation Coefficient," *The American Statistician*, 74:1, 29–36: Comment by Krishnamoorthy and Xia

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample from a bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}, N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Define

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

Hu, Jung, and Qin (2020) have considered the problem of interval estimating the correlation coefficient $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$, where σ_{ij} is the (i,j)th element of $\boldsymbol{\Sigma}$. In Section 3.1 of their article, the authors have developed generalized pivotal quantity (GPQ) using the generalized variable approach. To describe their GPQ, let S_{ij} denote the (i,j)th element of $\boldsymbol{S}$, and let s_{ij} be an observed value of S_{ij}. Let $s_{2.1} = s_{22} - s_{21}^2/s_{11}$. Furthermore, let Z be a standard normal random variable and U_i be a χ^2_{n-i} random variable, $i = 1, 2$. Assume that these random variables Z, U_1, and U_2 are independent. In terms of these random variables and observed values s_{ij}, the GPQ given in (3.9) of Hu, Jung, and Qin (2020) can be expressed as

$$G_\rho = \frac{\left(s_{21}/s_{11} - Z\sqrt{s_{2.1}/(s_{11}U_2)}\right)\sqrt{s_{11}/U_1}}{\left[\left(s_{21}/s_{11} - Z\sqrt{s_{2.1}/(s_{11}U_2)}\right)^2 s_{11}/U_1 + s_{2.1}/U_2\right]^{1/2}}. \tag{1}$$

For a given (s_{11}, s_{21}, s_{22}), the $100\alpha/2$ percentile and the $100(1-\alpha/2)$ percentile of G_ρ form a $100(1-\alpha)\%$ confidence interval (CI) for ρ. This CI is referred to as the generalized CI.

The above GPQ is not new and it has been already given in our article (Krishnamoorthy and Xia 2007). Indeed, we have obtained the GPQ for the correlation coefficient ρ as a special case from the GPQs of the elements of a $p \times p$ normal covariance matrix $\boldsymbol{\Sigma}$. The GPQs for the elements of $\boldsymbol{\Sigma}$ were obtained from the GPQs for the elements θ_{ij} of $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the Cholesky factor of $\boldsymbol{\Sigma}$. Using our GPQs, one can find CI for the difference between two overlapping correlation coefficients $\rho_{ij} - \rho_{ik}$ and for the difference between two non-overlapping correlation coefficients $\rho_{ij} - \rho_{kl}$, $i \neq j \neq k \neq l$. To write our GPQ for the simple correlation coefficient ρ, let $r = s_{21}/\sqrt{s_{11}s_{22}}$, the sample correlation coefficient, and let $r^* = r/\sqrt{1-r^2}$. Then our GPQ (see Krishnamoorthy and Xia 2007, eq. (16)) can be expressed as

$$Q_\rho = \frac{r^*\sqrt{U_2} - Z}{\sqrt{\left(r^*\sqrt{U_2} - Z\right)^2 + U_1}}. \tag{2}$$

For a given (s_{11}, s_{21}, s_{22}), the $100\alpha/2$ and $100(1-\alpha/2)$ percentiles of Q_ρ form a $(1-\alpha)$ confidence interval for ρ.

We can show that the GPQs G_ρ and Q_ρ are the same as follows. We can write the numerator of (1) as

$$\left(\frac{s_{21}}{s_{11}} - Z\sqrt{\frac{s_{2.1}}{s_{11}U_2}}\right)\sqrt{\frac{s_{11}}{U_1}} = \left(r^*\sqrt{U_2} - Z\right)\sqrt{\frac{s_{2.1}}{U_1 U_2}}.$$

Substituting the above expression in the numerator and in the denominator of (1), it can be readily verified that $G_\rho = Q_\rho$.

The simulation study by Krishnamoorthy and Xia (2007) indicated that the generalized CI for the correlation coefficient ρ is practically exact. Later, Krishnamoorthy (2013) has shown that the generalized CI is exact, and extended the results to the case of missing data.

References

Krishnamoorthy, K. (2013), "Comparison of Confidence Intervals for Correlation Coefficients Based on Incomplete Monotone Samples and Those Based on Listwise Deletion," *Journal of Multivariate Analysis*, 114, 378–388. [418]

Krishnamoorthy, K., and Xia, Y. (2007), "Inferences on Correlation Coefficients: One-Sample, Independent and Correlated Cases," *Journal of Statistical Planning and Inference*, 137, 2362–2379. [418]

Kalimuthu Krishnamoorthy
University of Louisiana at Lafayette
krishna@louisiana.edu

Yanping Xia
University of Nevada at Las Vegas
yanping.xia@unlv.edu