11-2016

# A Note on the Relationship between the Pearson Product-Moment and the Spearman Rank-Based Coefficients of Correlation

Todd C. Headrick
*Southern Illinois University Carbondale*, headrick@siu.edu

Open Journal of Statistics, 2016, 6, 1025-1027
http://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

# A Note on the Relationship between the Pearson Product-Moment and the Spearman Rank-Based Coefficients of Correlation

**Todd Christopher Headrick**

Department of CQMSE (Quantitative Methods-Statistics), Southern Illinois University, Carbondale, USA
Email: headrick@siu.edu

**How to cite this paper:** Headrick, T.C. (2016) A Note on the Relationship between the Pearson Product-Moment and the Spearman Rank-Based Coefficients of Correlation. *Open Journal of Statistics*, **6**, 1025-1027.
http://dx.doi.org/10.4236/ojs.2016.66082

**Received:** September 15, 2016
**Accepted:** November 14, 2016
**Published:** November 17, 2016



## Abstract

This note derives the relationship between the Pearson product-moment coefficient of correlation and the Spearman rank-based coefficient of correlation for the bivariate normal distribution. This new derivation shows the relationship between the two correlation coefficients through an infinite cosine series. A computationally efficient algorithm is also provided to estimate the relationship between the Pearson product-moment coefficient of correlation and the Spearman rank-based coefficient of correlation. The algorithm can be implemented with relative ease using current modern mathematical or statistical software programming languages e.g. R, SAS, Mathematica, Fortran, *et al.* The algorithm is also available from the author of this article.

## Keywords

Bivariate Normal Distribution, Product-Moment Correlation, Rank-Based Correlation, Gibbs Phenomenon

## 1. Introduction

The Pearson product-moment coefficient of correlation can be interpreted as the cosine of the angle between variable vectors in $n$ dimensional space (e.g. [1] and [[2], p. 702]). Pearson [3] showed that the relationship of turning Spearman rank-based correlation coefficients ( $\rho_S$ ) for the bivariate normal distribution into Pearson product-moment correlations ( $\rho$ ), which was contrived based on the so-called correlation of grades, for large samples to be:

$$\rho_S = (6/\pi)\sin^{-1}\left((1/2)\rho\right). \tag{1}$$

DOI: 10.4236/ojs.2016.66082 November 17, 2016

For finite (small) samples, Moran [4] derived the relationship between the Pearson and Spearman coefficients of correlation for the bivariate normal distribution, which also appears in Headrick [[5] p. 114], to be:

$$\rho_S = (6/\pi)\left\{((n-2)/(n+1))\sin^{-1}((1/2)\rho) + (1/(n+1))\sin^{-1}(\rho)\right\}. \tag{2}$$

Taking the limit as $n \to \infty$ in Equation (2) will reduce Equation (2) to Equation (1). We would also note that Höffding [6] demonstrated that the Spearman rank correlation tends to normality for any given parent population.

## 2. Mathematical Development

In view of the above, this note derives the relationship between the Pearson product-moment correlation coefficient and the Spearman rank-based correlation coefficient for the bivariate normal distribution, in a different manner from either the Pearson [3] or the Moran [4] derivations, through the following infinite cosine series:

$$\sum_{n=1}^{\infty} \cos(nx)/n. \tag{3}$$

Specifically, if we let $z = \cos(x) + i\sin(x)$, then

$$\sum_{n=1}^{m} y^{n-1} z^n = \left(z\left\{1-(yz)^m\right\}\right)/(1-yz) \tag{4}$$

where it follows that for $|y| < 1$, that

$$\begin{aligned} &\sum_{n=1}^{\infty} y^{n-1}(\cos(nx) + i\sin(nx)) \\ &= (\cos(x) + i\sin(x))/(1 - y\cos(x) - yi\sin(x)) \\ &= ((\cos(x) - y) + i\sin(x))/(1 - 2y\cos(x) + y^2). \end{aligned} \tag{5}$$

Thus, from Equation (5) we have:

$$\sum_{n=1}^{\infty} y^{n-1}\cos(nx) = (\cos(x) - y)/(1 - 2y\cos(x) + y^2). \tag{6}$$

The series associated with Equation (6) is uniformly convergent for all values of $y$ and for $|y| \le p < 1$. As such, integrating with respect to $y$, where $0 < y < 1$ yields:

$$\begin{aligned} \sum_{n=1}^{\infty} y^n(\cos(nx))/n &= \int_0^y ((\cos(x) - t)/(1 - 2t\cos(x) + t^2))\,\mathrm{d}t \\ &= (-1/2)(\ln(1 - 2y\cos(x) + y^2)). \end{aligned} \tag{7}$$

Let $x$ neither be zero nor a multiple of $2\pi$. As such, it necessarily follows that the series in Equation (3) is convergent. Hence, for $0 \le y \le 1$; $y^n$ is positive, monotonic, decreasing, and bounded. Whence, the series

$$\sum_{n=1}^{\infty} y^n \cos(nx)/n \tag{8}$$

is, therefore, uniformly convergent for $0 \le x \le 1$. Subsequently letting $y \to 1$, noting again that $x$ is neither zero nor a multiple of $2\pi$, it follows that Equation (3) can be expressed as

$$\sum_{n=1}^{\infty} \cos(nx)/n = (-1/2)\ln(2 - 2\cos(x)) = -\ln(2\sin((1/2)x)). \tag{9}$$

## 3. Main Result and Conclusions

Setting $x = (\pi/3)\rho_S$ in Equation (9), and through subsequent inverse exponentiation $(1/e)$ of Equation (9), yields the relationship (for large samples) between the Pearson product-moment correlation and the Spearman rank-based correlation coefficients as

$$\rho = 2\sin\left((\pi/6)\rho_S\right) \tag{10}$$

for the bivariate normal distribution. In conclusion, the algorithm provided below in Equation (11), which has an oscillating effect of the Gibbs phenomenon [7], to demonstrate the analytical derivation above is given as:

$$\hat{\rho} = \left(1 - (1/k)\sum_{n=1}^{k}\cos\left(n(\pi/3)\rho_S\right)/n\right)^k \tag{11}$$

where $0 \le |\rho_S| \le 1$, $k$ is finite, and where Equation (11) converges to Equation (10) as $k \to \infty$. Finally, in terms of the error associated with Equation (11), it is straight-forward to see through real analysis, that $\rho_S$ and $\rho$ have a maximum absolute deviation when $\rho_S = 0.566467\cdots$ and hence Equation (10) would result in $\rho = 0.584543\cdots$. As such, at this maximum point of deviation, given that $k = 10000$ in Equation (11), that the absolute error is less than $5.42 \times 10^{-5}$ when juxtaposed with Equation (10).

## References


[1] Rodgers, J.L. and Nicewander, W.A. (1988) Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, **42**, 59-66. https://doi.org/10.2307/2685263

[2] Stein, S.K. and Barcellos, A. (1992) Calculus and Analytic Geometry. 5th Edition, McGraw-Hill, Inc., New York.

[3] Pearson, K. (1907) Mathematical Contributions to the Theory of Evolution. XVI. On Further Methods of Determining Correlation. *Drapers Company of Research Memoirs*, *Biometric Series*, Cambridge University Press, Cambridge.

[4] Moran, P.A.P. (1948) Rank Correlation and Product-Moment Correlation. *Biometrika*, **35**, 203-206. https://doi.org/10.1093/biomet/35.1-2.203

[5] Headrick, T.C. (2010) Statistical Simulation: Power Method Polynomials and Other Transformations. Chapman & Hall/CRC, Boca Raton.

[6] Höffding, W. (1948) A Class of Statistics with Asymptotically Normal Distributions. *The Annals of Mathematical Statistics*, **19**, 293-325. https://doi.org/10.1214/aoms/1177730196

[7] Gibbs, J.W. (1899) Fourier Series. *Nature*, **59**, 200, 606.