On the Nonadditivity of Correlation Coefficients Part 3: On the Bias & Nature of Correlation Coefficients

Abstract – This is Part 3 of a tutorial series on the nonadditivity of correlation coefficients. The bias & nature of correlation coefficients, their transformations, and assumptions of normality are discussed.

Keywords: nonadditivity, normality, shifted up cosine approximation, regression, correlation, association, bias
Published: 11-01-2018; Updated: 11-11-2018

Note: This article is part of a legacy series that the author published circa 2011 at http://www.miislita.com, now a search engine site. It is published in pdf format here at http://www.minerazzi.com , with its content updated.

Introduction

This is the last part of a tutorial series on correlation coefficients. In Part 1 we discussed why we cannot arithmetically add, subtract, or average Pearson's Product-Moment Correlation Coefficients r_p, or Spearman's Rank-Order Correlation Coefficients r_s. In Part 2 we discussed that average correlation coefficients based on Fisher *r*-to-*Z* transformations are invalid for distributions that violate bivariate normality. (Garcia, 2017a; 2017b).

This time we want to discuss the so-called inherent bias of correlation coefficient estimates. We first discuss early work on the subject. After this, we investigate how this bias can be understood by deriving a connection between the following statistical concepts: regression, correlation, association, and determination. The notations adopted in Part 1 and 2 applies. For the sake of clarity, we also use the following naming conventions:

- r is either a Pearson or Spearman sample correlation.
- ρ is a population true correlation.
- $E[r]$ is the expected correlation of a random sample from a normal population. An expected value is also known as the average, mean, or first moment (Wikipedia, 2018).
- $|E[r] - \rho|$ is the absolute bias of the estimated correlation.
- n is the number of determinations or sample size.

Discussion

Ronald Aylmer Fisher found that the expected value of correlation coefficients based on random sampling from a population assumed normally distributed is approximately equal to (Fisher, 1915; Zimmerman, Zumbo, & Williams, 2003)

$$E[r] = \rho - \frac{\rho(1-\rho^2)}{2n} \quad (1)$$

where the absolute bias of the estimated correlation is

$$|E[r] - \rho| = \left|\frac{\rho(1-\rho^2)}{2n}\right| \quad (2)$$

We may refer to (2) as the Fisher bias. Olkin and Pratt (1958) suggested (3) for very small samples.

$$|E[r] - \rho| = \left|\frac{\rho(1-\rho^2)}{2n-3}\right| \quad (3)$$

which we may call the Olkin-Pratt bias.

Since (2) and (3) are zero when $\rho = 0$ or $\rho = 1$, one might think that the bias will reach a maximum at $\rho = 0.5$. Deriving (2) or (3) respect to ρ, setting the result equal to zero, and solving for ρ, the maximum bias actually occurs when $\rho_{max} = \sqrt{1/3} \approx 0.577$, where ρ_{max} is independent of *n*.

Substituting $\rho = \rho_{max}$ the maximum Fisher bias is $0.192/n$ and the maximum Olkin-Pratt bias is $0.385/(2n-3)$. Figure 1 shows the corresponding bias vs. ρ curves for *n* = 10. The bias is higher in the mid-range ($\rho = 0.5 - 0.7$) and the curves are skewed forward.

Figure 2 depicts the maximum bias for different sample sizes. When $n \geq 10$, the maximum Fisher bias is less than 0.020 and the Olkin-Pratt bias is less than 0.025. As *n* increases the maximum bias decreases and the two curves eventually agree.

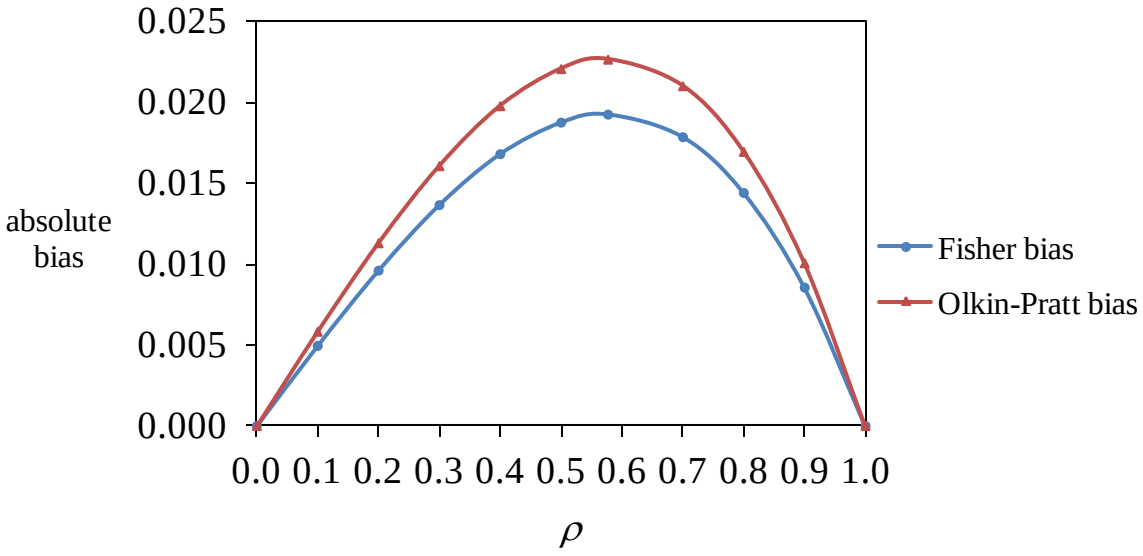

Figure 1. Absolute bias vs. ρ curves for n = 10.

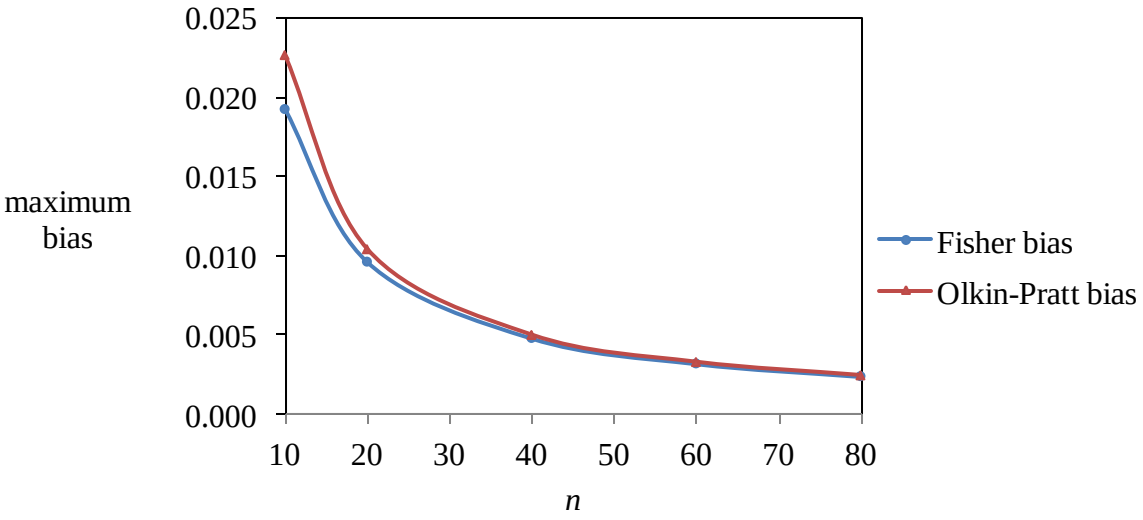

Figure. 2. Maximum bias as a function of the sample size.

Score-to-Rank Transformations

If the initial data are ranks, there is no difference between a Spearman and Pearson correlation coefficient because the only sampling distribution to be considered is due to ranks. However, if the initial data are scores, transforming these to ranks can change the sampling distribution of a correlation coefficient, introducing another source of bias. In this section we want to compare this bias with the one generated by (2) and (3).

It has been documented (Zimmerman et. al., 2003; Daniels, 1950, 1951; Durbin & Stuart, 1951; Höeffding, 1948a; 1948b; Kendall & Gibbons, 1990), that Spearman r_s and Pearson ρ can be related by the formula

$$E[r_s] = \frac{6}{\pi(n+1)}\left[\sin^{-1}\rho + (n-2)\sin^{-1}\frac{\rho}{2}\right] \quad (4)$$

where the absolute bias $|E[r] - \rho|$ due to score-to-rank transformations is

$$|E[r_s] - \rho| = \left|\frac{6}{\pi(n+1)}\left[\sin^{-1}\rho + (n-2)\sin^{-1}\frac{\rho}{2}\right] - \rho\right| \quad (5)$$

where for large samples (4) and (5) are approximately equal to

$$E[r_s] \approx \frac{6}{\pi}\left[\sin^{-1}\frac{\rho}{2}\right] \quad (6)$$

$$|E[r_s] - \rho| \approx \left|\frac{6}{\pi}\left[\sin^{-1}\frac{\rho}{2}\right] - \rho\right| \quad (7)$$

Deriving (5) respect to ρ, setting the result equal to zero, and solving for ρ, the maximum bias for large samples occurs at $\rho_{max} = 2\sqrt{1-(3/\pi)^2} \approx 0.594$

Figure 3 depicts bias vs. ρ profile curves for different values of n. As n increases, ρ approaches ρ_{max}. The bias is considerable in the mid-range ($\rho = 0.5 - 0.7$) for $n \leq 10$ and a lot less in the high-range ($\rho = 0.9 - 1.0$) for $n \geq 40$.

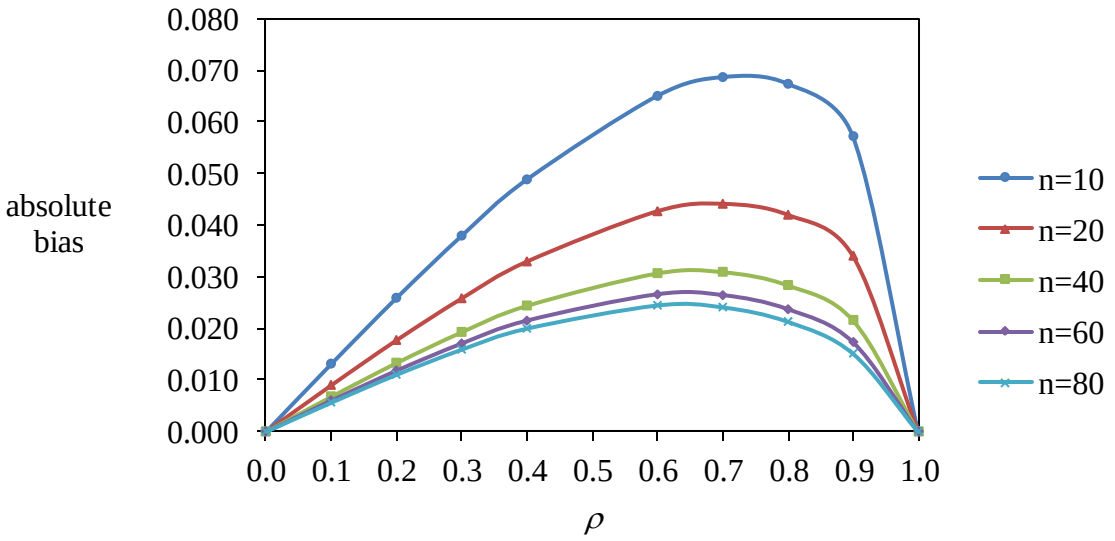

Figure 3. Absolute bias vs. ρ profile curves due to score-to-rank transformations.

Substituting $\rho = \rho_{max}$ in (5), we find that the maximum bias is less than 0.070 for n = 10 and less than 0.030 for n = 80. Figure 4 and Table 1 show a comparative of ρ vs. n curves as computed with (2), (3), and (5).

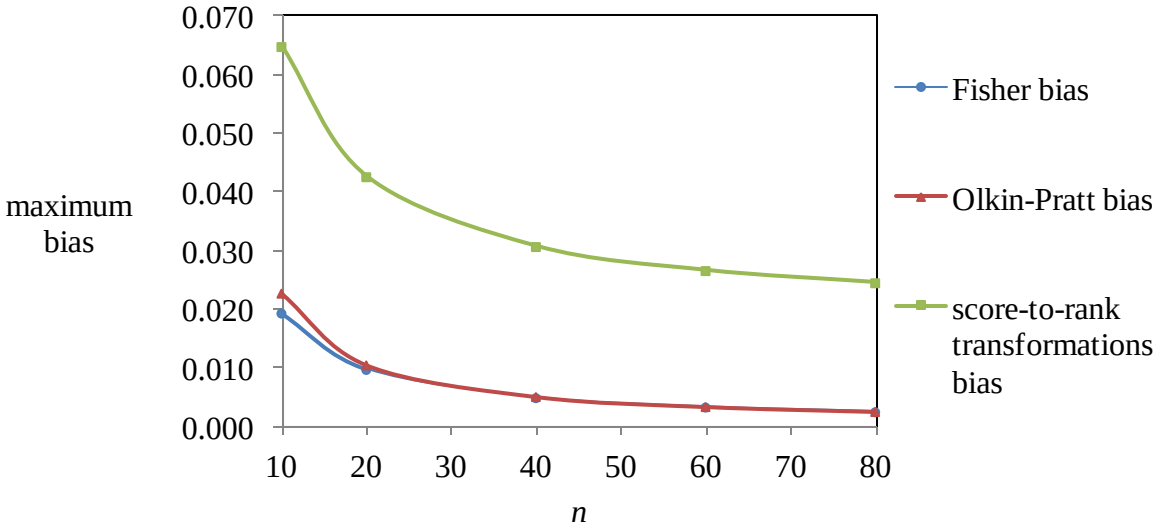

Figure 4. Comparative of maximum bias vs. *n* curves.

Table 1. Comparative of ρ vs. n data as computed with (2), (3), and (5).

ρ	n = 10	n = 20	n = 40	n = 60	n = 80
0.000	0.000	0.000	0.000	0.000	0.000
0.100	0.013	0.009	0.007	0.006	0.006
0.200	0.026	0.018	0.013	0.012	0.011
0.300	0.038	0.026	0.019	0.017	0.016
0.400	0.049	0.033	0.024	0.021	0.020
0.600	0.065	0.043	0.031	0.027	0.024
0.700	0.069	0.044	0.031	0.026	0.024
0.800	0.067	0.042	0.028	0.024	0.021
0.900	0.057	0.034	0.022	0.017	0.015
1.000	0.000	0.000	0.000	0.000	0.000
(5) $\rho_{\max} = 0.594$	0.065	0.043	0.031	0.026	0.024
(3) $\rho_{max} = 0.577$	0.023	0.010	0.005	0.003	0.002
(2) $\rho_{max} = 0.577$	0.019	0.010	0.005	0.003	0.002

Figure 4 and Table 1 reveal that the bias generated from score-to-rank transformations is by far larger than the Fisher and Olkin-Pratt bias. The last three rows of the table indicates that this is so by almost three times for n = 10 and ten times for n = 80. Furthermore, Zimmerman et. al. (2003) found that the expected correlations are less than ρ and even less when these transformations are applied.

Another source of bias, frequently overlooked, is due to the so-called Fisher transformations. These transformations are used for averaging correlation coefficients, significance testing, and computing confidence intervals and are discussed in the next section.

Fisher Transformations

Fisher (1915, 1921) proposed

$$Z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) \tag{8}$$

as a variance stabilizing transformation whose inverse is

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \frac{e^{Z} - e^{-Z}}{e^{Z} + e^{-Z}} \tag{9}$$

where (8) is called the *r-to-Z* transformation and (9) the *Z-to-r* transformations, and both are often referred to as the Fisher Transformations. Most programming languages list these as the inverse hyperbolic tangent (atanh) and hyperbolic tangent (tanh) functions. The transformations are also available in the calculator bundled with Windows ® as the $\tanh^{-1}$ and tanh functions, and in Excel ® as the ATANH and TANH functions.

Zimmerman, et. al. (2003) found that Fisher transformations should not be applied indiscriminately as they are not robust to non-normality. Indeed, these transformations should be applied when the data exhibit bivariative normality; i.e., when both variates, *x* and *y*, from which correlations are computed are approximately normal. The fact that score-to-rank transformations can change the sampling distribution of correlation coefficients should prevent researchers from blindly applying them. Nonnormality should never be ignored (Bishara & Hittner; 2015).

The sampling distribution of Z is normally distributed with standard error $S_e = 1/\sqrt{n-3}$ and confidence interval equal to $Z \pm zS_e$. Evidently the size of a confidence interval is $\Delta Z = 2zS_e$, independent of *r*, symmetrical around an *r-to-Z* transformed value, and known before any of its bounds are known.

The *z* term, not to be confused with the uppercase Z in (8), is a score at a given confidence interval that expresses the probability that a population statistic lies within that interval. These *z* scores are obtained from statistical tables.

For instance, a value of $z_{95\%} = 1.96$ corresponds to a confidence interval of 95%, and vice versa. The size and upper/lower bounds of said interval are:

$$\Delta Z = Z_{upper} - Z_{lower} = 2z_{95\%}S_e = \frac{2*1.96}{\sqrt{n-3}} = \frac{3.92}{\sqrt{n-3}} \quad (10)$$

$$Z_{upper} = Z + \frac{\Delta Z}{2} \quad (11)$$

$$Z_{lower} = Z - \frac{\Delta Z}{2} \quad (12)$$

Z-to-r transforming these bounds yields the r_{upper} and r_{lower} bounds of the confidence interval. The size of the confidence interval in the *r* scale is $\delta = r_{upper} - r_{lower}$ and, unlike ΔZ, depends of *r*, and is not symmetrical. We can estimate it without knowing its bounds by mapping ΔZ to the scale of *r* values, but a small error is introduced by doing so. Let us address this point.

Deriving (8) respect to *r*, it can be demonstrated that

$$\frac{dZ}{dr} = \frac{1}{(1-r)(1+r)} = \frac{1}{1-r^2} \quad (13)$$

where the $1 - r^2$ term is identified as the $\frac{dr}{dZ}$ derivative. Approximating $\frac{dr}{dZ}$ to a ratio of finite differences, the size of the confidence interval on the *Z* scale is mapped to the *r* scale as

$$\Delta r = (1 - r^2)\Delta Z \quad (14)$$

and the absolute error due to estimating the size of the confidence interval in this way is

$$|\Delta r - \delta| = \left|(1 - r^2)\Delta Z - r_{upper} + r_{lower}\right| \quad (15)$$

Table 2 lists results to three decimal places as computed from (8) through 15, for *n* = 34 and $z_{95\%} = 1.96$. Several conclusions can be drawn. First, changes of *Z* and *r* are nearly equal in the neighborhood of zero (Fisher, 1921). Second unlike ΔZ, Δr decreases as *r* increases. Third, a confidence interval on the *Z* scale is symmetrically distributed around an *r-to-Z* transformed value while a confidence interval on the *r* scale is not symmetrical around the corresponding *r* value because the distribution of *r* is skewed forward. Thus, *r* values are closer to r_{upper} than to r_{lower}.

Table 2 shows that the error introduced by (15), $|\Delta r - \delta|$, is less than 0.01 for correlations in the mid- and high-range (i.e., $r > 0.5$) and less than 0.03 for very small correlations ($r < 0.2$). The table also shows that for small samples the error reaches a minimum as r approaches 0.6.

Table 2. Confidence intervals data (n = 34) on the Z and r scales at the 95% confidence level.

r	Z	Z_{upper}	Z_{lower}	ΔZ	r_{upper}	r_{lower}	δr	Δr	$\lvert\Delta r - \delta\rvert$
0.050	0.050	0.402	-0.302	0.704	0.382	-0.293	0.675	0.702	0.027
0.100	0.100	0.452	-0.252	0.704	0.424	-0.247	0.670	0.697	0.027
0.200	0.203	0.555	-0.149	0.704	0.504	-0.148	0.652	0.676	0.024
0.300	0.310	0.662	-0.043	0.704	0.579	-0.042	0.622	0.641	0.019
0.400	0.424	0.776	0.072	0.704	0.650	0.072	0.579	0.591	0.013
0.500	0.549	0.901	0.197	0.704	0.717	0.195	0.522	0.528	0.006
0.600	0.693	1.045	0.341	0.704	0.780	0.328	0.451	0.451	0.001
0.700	0.867	1.219	0.515	0.704	0.839	0.474	0.365	0.359	0.006
0.800	1.099	1.451	0.747	0.704	0.896	0.633	0.263	0.253	0.009
0.900	1.472	1.824	1.120	0.704	0.949	0.808	0.142	0.134	0.008
0.950	1.832	2.184	1.480	0.704	0.975	0.901	0.074	0.069	0.005

Figure 5 shows profile curves of $|\Delta r - \delta|$ vs. r for different sample sizes. Essentially, when n increases the error decreases and the minimum r tends to vanish. Therefore for large samples, we can fairly predict one bound on the r scale if we know the other bound; for instance if we know r_{upper} then r_{lower} can be calculated as

$$r_{lower} = r_{upper} - \Delta r \tag{16}$$

So instead of computing r_{lower} with a *Z-to-r* transformation, we may apply (16), which is less computationally expensive.

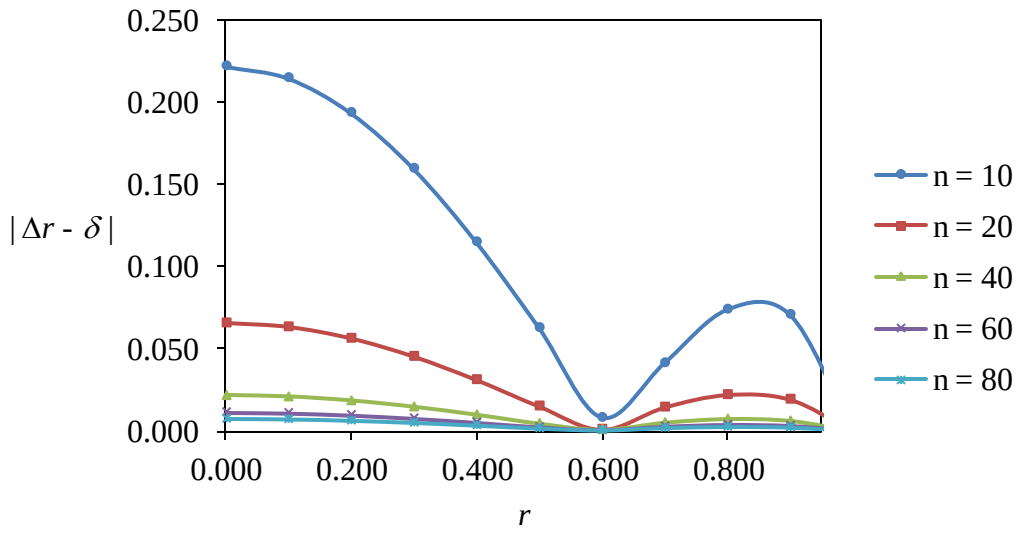

Figure 5. Profile curves of $|\Delta r - \delta|$ vs. r for different sample sizes.

All of the above sources of bias seem to have a common denominator: the normality assumption.

The Normality Assumption as a Common Source of Bias

A normal distribution is one with mean equal to 0 and variance equal to 1. For correlation coefficients, this normality assumption is applied at two different levels by assuming that:

(a) the sampling distribution of a correlation coefficient is normal.
(b) the x and y variables from which a correlation coefficient is computed are normally distributed (bivariate normality).

The Fisher (2) or Olkin-Pratt (3) bias is a property of the mean (Zimmerman, et. al., 2003) which is derived assuming (a). On the other hand, the bias due to score-to-rank transformations is also the result of adopting (a). We must keep in mind that these transformation can change the sampling distribution of a correlation coefficient. Even if this is not the case, computing an r_s as an r_p is not a license to claim normality since an r_s requires no assumptions about normality or linearity.

Regarding the bias due to Fisher Transformations, let's recall that these are sensitive to bivariative normality violations. If one or both of the variates lack of normality, relying on these transformations can lead to misleading significance tests, confidence intervals, and correlation averages, particularly if the size of a confidence interval is comparable with the bias generated.

Furthermore, blindly applying Fisher Transformations to correlations that are the result of converting scores to ranks simply compounds the bias from said conversions. The obvious conclusion is that the above different sources of bias have a common denominator: the assumption of normality. This assumption is implicit in almost all previous research on correlation theory (Hotelling, 1953; Moran, 1948; Gayen, 1951; Hawkings, 1989; Bishara & Hittner, 2015). What is left to check is if the actual nature of correlation coefficients conforms to the assumption.

On the Cosine Nature of Correlation Coefficients

If we apply simple linear regression to a set of paired scores (x, y), the following model is obtained

$$y = \beta_0 + \beta_1 x \tag{17}$$

where β_0 is the intercept and β_1 the slope of the regression model. It can be shown that

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{18}$$

$$\beta_1 = r_p \left(\frac{s_y}{s_x}\right) \tag{19}$$

$$r_p = \frac{covar(x,y)}{s_x s_y} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (y_i - \bar{y})^2}} = \frac{\sum_i^n (x_i')(y_i')}{\sqrt{\sum_i^n (x_i')^2}\sqrt{\sum_i^n (y_i')^2}} \tag{20}$$

$$r_p = \frac{\sum_i^n (x_i')(y_i')}{\sqrt{\sum_i^n (x_i')^2}\sqrt{\sum_i^n (y_i')^2}} = \cos\theta = \frac{\vec{x} \bullet \vec{y}}{\|\vec{x}\| \, \|\vec{y}\|} \tag{21}$$

$$r_p^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2} = (\cos\theta)^2 = \frac{1 + \cos 2\theta}{2} \tag{22}$$

The above expressions are really enlightening. First, notice from (17) to (19) that as a concept, regression subsumes correlation via the β_1 coefficient: We may compute r_p from the slope of a simple linear regression model as $\left(\frac{s_x}{s_y}\right)\beta_1$, where s_x and s_y are standard deviations.

Second, if in (18) both means are not zero ($\bar{x} \neq 0,\ \bar{y} \neq 0$), r_p is based on mean-centering scores, as can be seen from (20) since $x_i' = x_i - \bar{x}$ and $y_i' = y_i - \bar{y}$.

Third, notice from (21) that r_p is a cosine: the cosine similarity between two vectors (Cook, 2010a, 2010b; 2010c; Fisher, 1924; Garcia, 2017a). Thus, a centered cosine similarity is a centered correlation coefficient and vice versa. Both are non-additive measures of associations.

Because in (22) $\hat{y}_i$ is a regressed value, it follows that $\sum_i^n (\hat{y}_i - \bar{y})^2$ is a regression sum of squares and $\sum_i^n (y_i - \bar{y})^2$ a total sum of squares. Notice that the coefficient of determination r_p^2 is a cosine-squared function.

By contrast, a zero-intercept regression ($\beta_0 = 0$, also known as no-intercept regression and as regression through zero or through origin) requires that $\bar{x} = \bar{y} = 0$ which leads to the following uncentered statistics denoted by a zero subscript, where υ is the number of degrees of freedom.

$$covar(x,y)_0 = \frac{\sum_i^n (x_i)(y_i)}{\upsilon} \tag{23}$$

$$s_{x_0} = \sqrt{\frac{\sum_i^n (x_i)^2}{\upsilon}} \tag{24}$$

$$s_{y_0} = \sqrt{\frac{\sum_i^n (x_i)^2}{\upsilon}} \tag{25}$$

$$r_{p_0} = \frac{covar\,(x,y)_0}{s_{x_0} s_{y_0}} = \frac{\sum_i^n (x_i)(y_i)}{\sqrt{\sum_i^n (x_i)^2}\sqrt{\sum_i^n (y_i)^2}} = \cos\theta_0 \tag{26}$$

$$r_{p_0}^2 = \frac{\sum_i^n (\hat{y}_i)^2}{\sum_i^n (y_i)^2} = (\cos\theta_0)^2 = \frac{1+\cos 2\theta_0}{2} \tag{27}$$

Note from (26) that r_{p_0} is an uncentered correlation coefficient which in turns is the uncentered cosine similarity that is commonly computed from two data sets treated as vectors. Again, both are association measures.

Last, but not least, the numerator and denominator of (27) represent the sum of squares around zero while the numerator and denominator of (22) represent the sum of square around the mean of the dependent variable ($\bar{y}$).

Thus, when fitting data to a zero-intercept regression model, the data should be interpreted using uncentered statistics; otherwise we may be drawing misleading conclusions. In this regard, comparing centered and uncentered statistics like coefficients of determination, correlation coefficients, and so forth is like "comparing apples with oranges". Therefore, one cannot state that an uncentered regression model is better or worse than the centered one, for instance, by just comparing coefficients of determinations. (Kozac & Kozac, 1995).

Figure 6 summarizes (17) through (22) for the corresponding mean-centered statistics. Similar statistics can be derived from the uncentered case by setting mean values to zero.

$$\boldsymbol{(x,y)\ scores}\ \}\quad \boldsymbol{Data}\quad \{\ \boldsymbol{(R_x,R_y)\ ranks}$$

$$\left.\begin{array}{l} y=\beta_0+\beta_1 x \\ \beta_0=\bar{y}-\beta_1\bar{x} \\ \beta_1=r_p\left(\frac{s_y}{s_x}\right) \end{array}\right\}\ Regression\ \begin{cases} R_y=\beta_0+\beta_1 R_x \\ \beta_0=\overline{R_y}-\beta_1\overline{R_x} \\ \beta_1=r_s\left(\frac{s_{R_y}}{s_{R_x}}\right) \end{cases}$$

$$\left.\begin{array}{c} r_p=\dfrac{covar(x,y)}{s_x s_y} \\ Pearson \end{array}\right\}\ Correlation\ \begin{cases} r_s=\dfrac{covar(R_x,R_y)}{s_{R_x}s_{R_y}} \\ Spearman \end{cases}$$

$$\left.\begin{array}{c} r_p=\cos\theta \\ \cos\theta=\dfrac{\vec{x}\bullet\vec{y}}{\|\vec{x}\|\,\|\vec{y}\|} \\ cosine\ similarity \end{array}\right\}\ Association\ \begin{cases} r_s=\cos\theta \\ \cos\theta=\dfrac{\overrightarrow{R_x}\bullet\overrightarrow{R_y}}{\|\overrightarrow{R_x}\|\,\|\overrightarrow{R_y}\|} \\ cosine\ similarity \end{cases}$$

$$\left.\begin{array}{l} r_p^2=\dfrac{\sum_i^n\left(\hat{y}_i-\bar{y}\right)^2}{\sum_i^n\left(y_i-\bar{y}\right)^2}=(\cos\theta)^2 \\ r_p^2=(\cos\theta)^2=\dfrac{1+\cos 2\theta}{2} \end{array}\right\}\ Determination\ \begin{cases} r_s^2=\dfrac{\sum_i^n\left(\widehat{R_{y_i}}-\overline{R_y}\right)^2}{\sum_i^n\left(R_{y_i}-\overline{R_y}\right)^2}=(\cos\theta)^2 \\ r_s^2=(\cos\theta)^2=\dfrac{1+\cos 2\theta}{2} \end{cases}$$

Figure 6. Relationships between several statistical concepts.

Figure 6 shows that correlation coefficients are cosines while coefficients of determinations are cosines shifted up by 1 and scaled by a factor of 2. In both cases, these are even periodic functions and far from describing a bell-shaped curve. However, we may construct a nonperiodic representation by considering a one period of a cosine and check if said representation provides an approximation to normality.

Thus, what we need is to find a model that let's us do that.

Cosine Approximations to Normality

Back in 1961, Raab & Green proposed a shifted-up cosine function in the range [-π, π] as an approximation of normal distribution so that one period of a cosine has a probability density function (PDF)

$$f(x) = \frac{1+\cos x}{2\pi} \tag{28}$$

between [-π, π] and zero outside. Figure 7 depicts its PDF which can be generated with Excel ® or a software library (Heckert, A. & Filliben; 1997).

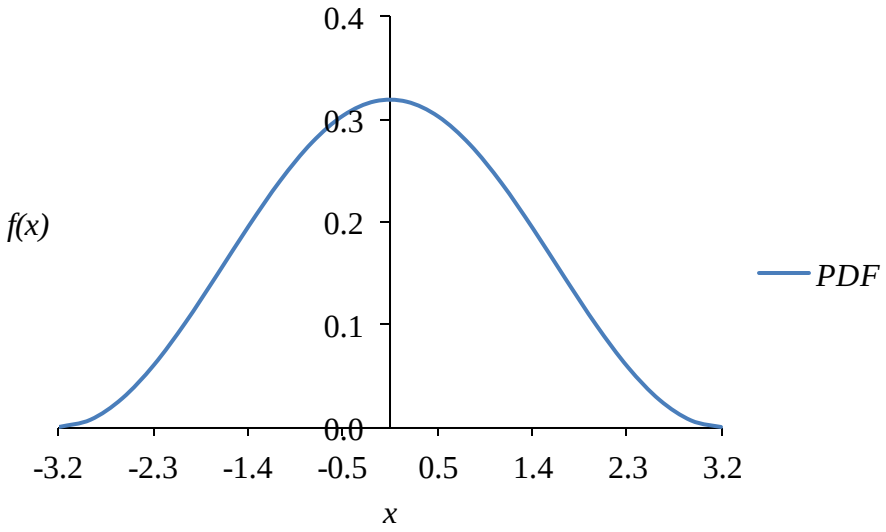

Figure 7. PDF of a cosine function between [-π, π] and generated with Excel ®.

The figure shows that (28) describes a bell-shaped curve (Raab & Green, 1961; Heckert & Filliben, 1997) with a maximum at $f(x=0) = 1/\pi \approx 0.32$, mean at x = 0,

$$\int_{-\pi}^{\pi} x f(x) dx = 0 \tag{29}$$

and variance (in radians) equal to

$$\int_{-\pi}^{\pi} x^2 f(x) dx = \frac{\pi^2}{3} - 2 \approx 1.29 \tag{30}$$

Notice that assuming normality requires a mean of zero and variance equal to 1, but (30) deviates considerably so the approximation is not that great (Cook, 2010a; 2010b; 2010c). These results suggest that taking at face value the distribution of a cosine (e.g., a correlation) for a normal distribution (mean = 0, variance = 1) might be the root of all sources of bias.

Raab & Green arbitrarily proposed (28) without providing any guidance on how to optimize the cosine approximation to normality. That may be why their proposal was not further developed in statistical publications. Warsza, Korczynski, and Galovska (2009) improved Raab & Green proposal by applying a shifted up cosine approximation model with optimized parameters. This model is based on a PDF of the general form

$$f(x) = B + A \cos 2\pi \frac{x}{X_T} \tag{31}$$

In (31), x is the value of observations, $f(x) > 0$, X_T is the range equal to one period of a cosine, and A (the amplitude) and B (a shifted up term) are constant parameters to be optimized. Because of the constant parameter B, the authors called (31) a *shifted up cosine* function and denoted it by the symbol $+COS$.

Different distribution approximations are then possible by fine tuning A and B. The authors found that a $+COS$ with $A = B$,

$$f(x) = A\left(1 + \cos 2\pi \frac{x}{X_T}\right) \tag{32}$$

and called COS^2 by the authors, can fairly approximate several distributions. They demonstrated that an optimized COS^2 can even approximate the Normal Distribution in the range ±2.5 standard deviations with accuracy of about ±0.02.

An optimized COS^2 with $A = B = \frac{1}{2}$ can then be applied to approximate to normality the PDF of a correlation coefficient, which is a cosine. Since (22) and (32) are of similar form; i.e.,

$$r_p^2 = \frac{1+\cos 2\theta}{2} = \frac{1}{2}(1 + \cos 2\theta) \tag{33}$$

a similar treatment can be applied to the definition of a coefficient of determination itself.

Additional research with real data is necessary to assess if optimized COS^2 functions effectively minimizes the bias of correlations coefficients and how effective they are in comparison to other methods (Bishara & Hittner; 2015).

Conclusion

The assumption of normality for the sampling distribution of correlation coefficients generates a bias that is a property of the mean, herein called the Fisher bias (2) and Olkin-Pratt bias (3).

Score-to-rank transformations generate a bias larger than the Fisher and Olkin-Pratt bias. Blindly applying *r-to-Z* transformations to score-to-rank transformed data is a recipe for a statistical disaster and should be avoided; otherwise we might end compounding the several sources of bias herein discussed.

Shifted up cosine approximations fairly approximates PDFs to normality, but additional research is necessary along these lines.

References

Bishara, A. J. & Hittner, J. B. (2015). Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality. *Educational and Psychological Measurement*. 75(5), pp 785-804. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5965513/ See also http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC5965513&blobtype=pdf

Cook, J. D. (2010a). Simple Approximation to Normal Distribution. Retrieved from https://www.johndcook.com/blog/2010/04/29/simple-approximation-to-normal-distribution/

Cook, J. D. (2010b). Cosines and Correlations. Retrieved from https://www.johndcook.com/blog/2010/06/17/covariance-and-law-of-cosines/

Cook, J. D. (2010c). Always Invert. Retrieved from https://www.johndcook.com/blog/2010/05/05/always-invert/

Daniels, H. E. (1950). Rank Correlation and Population Models. *Journal of the Royal Statistical Society*. Series B. 12, pp 171-181.

Daniels, H. E. (1951). Note on Durbin and Stuart's Formula for $E(r_s)$. *Journal of the Royal Statistical Society*. Series B, 13, 310.

Durbin, J. & Stuart, A. (1951). Inversions of rank correlation coefficients. *Journal of the Royal Statistical Society*. Series B, 13, pp 303-309.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*. 10, pp 507-521. Retrieved from http://digital.library.adelaide.edu.au/dspace/bitstream/2440/15166/1/4.pdf

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*. 1, pp 3-31. Retrieved from https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15169/1/14.pdf

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron*. 3, pp 329-332. Retrieved from https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15182/1/35.pdf

Garcia, E. (2017a). On the Nonadditivity of Correlation Coefficients Part 1: Pearson's *r* and Spearman's r_s. Retrieved from http://www.minerazzi.com/tutorials/nonadditivity-correlations-part-1.pdf

Garcia, E. (2017b). On the Nonadditivity of Correlation Coefficients Part 2: Fisher Transformations. Retrieved from http://www.minerazzi.com/tutorials/nonadditivity-correlations-part-2.pdf

Gayen, A. K. (1951). The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Non-Normal Universes. *Biometrika*. 38 (1/2), pp 219-247. JSTOR 2332329. doi:10.1093/biomet/38.1-2.219. Retrieved from https://academic.oup.com/biomet/article-abstract/38/1-2/219/245776/THE-FREQUENCY-DISTRIBUTION-OF-THE-PRODUCT-MOMENT?redirectedFrom=fulltext

Hawkins, D. L. (1989). Using U statistics to derive the asymptotic distribution of Fisher's Z statistic. *The American Statistician*. 43 (4): 235–237. American Statistical Association. JSTOR 2685369. doi:10.2307/2685369. Retrieved from https://www.jstor.org/stable/2685369?seq=1#page_scan_tab_contents

Heckert, A. & Filliben, J. J. (1997). COSPDF. *Dataplot Reference Manual*. Chapter 8: Probability Library Functions. Retrieved from https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cospdf.pdf

Höeffding, W. (1948a). A Non-Parametric Test of Independence. *The Annals of Mathematical Statistics*. Vol. 19, pp 546-557. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730150

Höeffding, W. (1948b). A Class of Statistics with Asymptotically Normal Distributions. *The Annals of Mathematical Statistics*. Vol. 19(3), pp 293-325. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730196

Hotelling, H (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society*. Series B. Blackwell Publishing. 15 (2): 193–225. JSTOR 2983768. Retrieved from https://www.jstor.org/stable/2983768?seq=1#page_scan_tab_contents

Kendall, M. G. & Gibbons, J. D. (1990). *Rank Correlation Methods*. (5th edition). Oxford University Press.

Kozak, A. & Kozak, R. A. (1995). Notes on regression through the origin. *The Forestry Chronicle*. MAI/JUIN Vol. 71, No. 3. Retrieved from http://pubs.cif-ifc.org/doi/pdf/10.5558/tfc71326-3

Moran, P. A. P. (1948). Rank Correlation and Product-Moment Correlation *Biometrika*. 35, (1-2), pp 203-206. Retrieved from https://doi.org/10.1093/biomet/35.1-2.203

Olkin, I. & Pratt, J. W. (1958). Unbiased Estimation of Certain Correlation Coefficients. Annals of Mathematical Statistics. 29, pp 201-211. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.aoms/1177706717

Raab, D. H. & Green, E. H. (1961). A Cosine Approximation to the Normal Distribution. *Psychometrika*. Vol 26, No. 4, pp 447. Retrieved from https://link.springer.com/article/10.1007/BF02289774

Warsza, Z. L., Korczynski, M. & Galovska, M. (2009). Shifted up Cosine Function as Model of Probability Distribution. *IMEKO World Congress*. pp 2417-2422. Retrieved from https://www.researchgate.net/publication/242173262_SHIFTED_UP_COSINE_FUNCTION_AS_MODEL_OF_PROBABILITY_DISTRIBUTION?enrichId=rgreq-26f940575714e7717979d86661aa682f-XXX&enrichSource=Y292ZXJQYWdlOzI0MjE3MzI2MjtBUzoxODU3MTk1OTEwMjI1OTJAMTQyMTI5MDQwOTA5Nw%3D%3D&el=1_x_2&_esc=publicationCoverPdf

Wikipedia (2018). Expected value. Retrieved from https://en.wikipedia.org/wiki/Expected_value

Zimmerman, D. W., Zumbo, B. D. & Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicológica*. 24:133–158. Retrieved from http://www.uv.es/revispsi/articulos1.03/9.ZUMBO.pdf See also http://www.redalyc.org/articulo.oa?id=16924109