Classication of bivariate measures of linear association

Elem. Math. 65 (2010) 1 – 11
0013-6018/10/010001-11
DOI 10.4171/EM/132

Elemente der Mathematik

Classification of bivariate measures of linear association

Wiebe R. Pestman

Wiebe R. Pestman obtained his doctor's degree in mathematics at the University of Groningen. He is currently a lecturer at the University of Utrecht. His mathematical interests are in functional analysis, probability and statistics, operator algebras, and harmonic analysis.

1 Introduction

Given an $n \times 2$ matrix, its rows can be geometrically interpreted as a cloud of points in the plane. Such a cloud of points could show clustering around a straight line. In statistics Pearson's (empiric) correlation coefficient is often used to capture the degree to which this phenomenon occurs in a number. This seems a peculiar usage, for Pearson's correlation coefficient changes when the cloud of points is rotated. Hence, whereas the shape of the cloud of points is completely preserved, the very number that is supposed to characterize clustering around a line changes its values (see also [4], [5], [6]). Other measures of linear association that do better in this respect can easily be created. Such measures, however, fail as a rule to be invariant under the action of rescaling the data. It will be explained that this is no coincidence.

2 Setting notations and conventions

The mean of the components $x_1, x_2, \ldots, x_n$ of a vector $\mathbf{x}$ will be denoted by $\bar{x}$. The covariance of two n-vectors $\mathbf{x}$ and $\mathbf{y}$ will be denoted by $\mathrm{cov}(\mathbf{x}, \mathbf{y})$, hence one has

$$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

In dem nachfolgenden Beitrag geht es darum, Maße zu finden, welche die Anpassung einer Punktemenge der euklidischen Ebene an eine Gerade beschreiben. Ein solches Maß ist beispielsweise der auf Pearson zurückgehende Korrelationskoeffizient, der sensitiv für ein solches „Clustering" von Punkten ist. Der Autor stellt nun fest, dass das Pearsonsche Maß nicht invariant gegenüber Drehungen, wohl aber gegenüber Streckungen in der x- bzw. y-Richtung ist. Insbesondere beweist er den Satz, dass es kein solches Maß gibt, dass sowohl gegenüber Rotationen als auch Skalierungen invariant ist.

The variance of a vector $\mathbf{x}$ is by definition the quantity $\mathrm{cov}(\mathbf{x}, \mathbf{x})$ and will be denoted by $\mathrm{var}(\mathbf{x})$. Pearson's correlation coefficient for two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\rho_P(\mathbf{x}, \mathbf{y}) = \frac{\mathrm{cov}(\mathbf{x}, \mathbf{y})}{\sqrt{\mathrm{var}(\mathbf{x})}\sqrt{\mathrm{var}(\mathbf{y})}}$$

provided the denominator in the fraction above is non-zero.

To every cloud of points, consisting of n points, belongs an $n \times 2$ matrix and vice versa. If the cloud of points contains at least two different points, then its corresponding matrix will be called a *datamatrix*. A datamatrix will be called *degenerate* if the corresponding points in the plane are collinear. Datamatrices will, as a rule, be denoted by the symbol $\mathbf{d}$. The first and second column of an $n \times 2$ matrix will sometimes be denoted by $\mathbf{x}$ and $\mathbf{y}$. These columns can be interpreted as vectors in $\mathbb{R}^n$. The Pearson correlation coefficient of a datamatrix $\mathbf{d}$ is understood to be the coefficient of its two columns and will be denoted as $\rho_P(\mathbf{d})$. In cases where there is one constant column the coefficient is set to 1 by convention.

3 Transformation groups and invariance

Suppose an $n \times 2$ datamatrix $\mathbf{d}$ is given. When rotating the corresponding cloud of points over an angle φ the matrix belonging to the rotated cloud is given by

$$\mathbf{d}' = \mathbf{d}\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}.$$

The fact that Pearson's correlation coefficient changes under rotation could thus mathematically be expressed by saying that one generally has

$$\rho_P(\mathbf{d}') \neq \rho_P(\mathbf{d}).$$

A 2×2 matrix of type

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

will be called a *rotation*. The set of all rotations forms a so-called *matrix multiplication group*, that is to say (see [1], [3]):

- Every rotation has an inverse and the inverse matrix is a rotation.
- The matrix product of two arbitrary rotations is a rotation.

In the following this group will be referred to as the *group of rotations*. The fact that Pearson's correlation coefficient changes its value under rotations will be referred to by saying that the coefficient is not invariant under the group of rotations.

Similar to rotating a datamatrix one can *scale* an $n \times 2$ matrix $\mathbf{d}$ to obtain a transformed datamatrix $\mathbf{d}'$ given by

$$\mathbf{d}' = \mathbf{d}\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

where $\alpha, \beta > 0$. The set of all matrices of type

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

with $\alpha, \beta > 0$ is also a matrix group. It will be called the *diagonal group*. Note that Pearson's correlation coefficient is invariant under this diagonal group.

A *translation* of a datamatrix $\mathbf{d}$ is understood to be a transformation of the form

$$\mathbf{d}' = \mathbf{d} + \mathbf{a}$$

where $\mathbf{a}$ is a matrix of type

$$\begin{pmatrix} p & q \\ p & q \\ \vdots & \vdots \\ p & q \end{pmatrix}.$$

Note that Pearson's coefficient is invariant under translations.

4 General measures of linear association

In this section a small framework will be set up in an effort to capture the concept of a bivariate measure of linear association in general terms. Such measures will be defined as being functions of a certain type, defined on the set of all datamatrices. The functions will have to meet certain continuity conditions. To be more precise in this, denote the set of all 2-column datamatrices with k rows by $\mathbb{D}_k$. The set $\mathbb{D}$ is now defined as the union of the $\mathbb{D}_1, \mathbb{D}_2, \ldots$ So $\mathbb{D}$ comprises all possible 2-column datamatrices, indifferent their number of rows.

Definition 1. A real-valued function $\mathbf{d} \mapsto \gamma(\mathbf{d})$, defined on the set of all 2-column datamatrices $\mathbb{D}$, is said to be *continuous* at the point $\mathbf{d}$ if for every sequence $\mathbf{d}_1, \mathbf{d}_2, \ldots$ of matrices (of the same size) converging to $\mathbf{d}$ the sequence $\gamma(\mathbf{d}_1), \gamma(\mathbf{d}_2), \ldots$ converges to $\gamma(\mathbf{d})$.

Definition 2. A bivariate measure of linear association is understood to be a real-valued function $\mathbf{d} \mapsto \gamma(\mathbf{d})$, defined on the set of all 2-column datamatrices $\mathbb{D}$, that has the following six properties:

a) For every $\mathbf{d}$ one has $0 \leq \gamma(\mathbf{d}) \leq 1$.

b) One has $\gamma(\mathbf{d}) = 1$ if and only if $\mathbf{d}$ is degenerate.

c) γ is invariant under row permutations.

d) γ is invariant under translations.

e) γ is invariant under scalar multiplication, that is to say, one has $\gamma(\mathbf{d}) = \gamma(s\,\mathbf{d})$ for every non-zero scalar s.

f) If a datamatrix $\mathbf{d}$ does not contain columns that are constant then γ is continuous at $\mathbf{d}$.

A measure will be called *symmetric* if $\gamma(\mathbf{d})$ remains unchanged when exchanging the two columns in $\mathbf{d}$.

The absolute value of Pearson's correlation coefficient could serve as a first example of a (symmetric) measure of linear association.

Definition 3. If a measure of linear association γ is invariant under rotations, then it will be called a *geometric* measure. Similarly, if γ is invariant under the diagonal group, then it will be called an *algebraic* measure.

General measures always being invariant under translations, geometric measures are actually invariant under the group of Euclidean transformations. For this reason they present geometric characteristics in the sense of modern geometry (see [9]). The absolute value of Pearson's coefficient, however, is algebraic, not geometric. It does not present a geometric characteristic. Clustering around a line being a geometric property, Pearson's coefficient seems not the right measure to capture this phenomenon. Following lines in [5], [4], here is an example of a geometric measure:

Definition 4. Given an arbitrary $n \times 2$ matrix $\mathbf{d}$ its *Euclidean correlation coefficient* is given by

$$\rho_E(\mathbf{d}) = \max |\rho_P(\mathbf{d}')|$$

where the maximum is taken over all rotated datamatrices $\mathbf{d}'$ of $\mathbf{d}$.

More colloquially, the Euclidean correlation coefficient is the maximum value of Pearson's correlation coefficient that can be obtained by rotating the cloud of points. Thus the properties listed in Definition 2 are met and therefore $\mathbf{d} \mapsto \rho_E(\mathbf{d})$ is a measure of linear association indeed. Besides this, exploiting the fact that rotations form a group, it is easy to see that the Euclidean correlation coefficient is invariant under rotations. However, the Euclidean correlation coefficient fails to be invariant under scalings. As will become apparent in the next theorem, this is no coincidence.

Theorem 5. *A measure of linear association cannot possibly be both geometric and algebraic.*

Proof. Suppose that $\mathbf{d} \mapsto \gamma(\mathbf{d})$ is a measure of linear association that is both invariant under the rotation and the diagonal group. Then it is also invariant under transformations of type

$$\mathbf{q_1 s q_2}$$

with $\mathbf{q_1}, \mathbf{q_2}$ rotations and $\mathbf{s}$ a positive diagonal matrix. However, every invertible 2×2 matrix can be decomposed in this way (see for example [10]). It follows from this that γ is invariant under the group of all invertible 2×2 matrices. Otherwise stated, for an arbitrary datamatrix and an arbitrary invertible 2×2 matrix $\mathbf{g}$ one would have

$$\gamma(\mathbf{d}) = \gamma(\mathbf{dg}).$$

It will turn out that this is impossible. To see this, define for every straight line ℓ in $\mathbb{R}^2$ the matrix $\mathbf{p}(\ell)$ as the 2×2 matrix belonging to the orthogonal projection on ℓ. Now fix an

arbitrary non-degenerate datamatrix $\mathbf{d}$ and choose ℓ in such a way that the matrix product $\mathbf{dp}(\ell)$ does not contain constant columns. Define the sequence $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 \ldots$ by

$$\mathbf{g}_\nu = \mathbf{p}(\ell) + \frac{1}{\nu}\big(\mathbf{e} - \mathbf{p}(\ell)\big)$$

where $\mathbf{e}$ stands for the identity matrix. Then $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 \ldots$ is a sequence of invertible matrices that converges to the non-invertible matrix $\mathbf{p}(\ell)$. Because $\mathbf{dp}(\ell)$ does not contain constant columns the measure γ is continuous there and for that reason one may write

$$1 = \gamma(\mathbf{dp}(\ell)) = \lim_{\nu\to\infty} \gamma(\mathbf{dg}_\nu) = \lim_{\nu\to\infty} \gamma(\mathbf{d}) = \gamma(\mathbf{d}).$$

The value 1, however, is exclusively reserved for degenerate matrices. Thus the assumption that γ is both geometric and algebraic leads to a contradiction. □

5 Maximizing Pearson's coefficient to obtain the Euclidean

In this section the main result in [5], [4] will be derived by using linear algebra. The problem that will be dealt with is the maximization of the expression

$$|\,\rho_P(\mathbf{dq})|$$

where $\mathbf{q}$ runs through the group of all rotations. Denoting the first and the second column of $\mathbf{d}$ by $\mathbf{x}$ and $\mathbf{y}$, the covariance matrix $\mathbf{C}(\mathbf{d})$ of $\mathbf{d}$ is defined as

$$\mathbf{C}(\mathbf{d}) = \begin{pmatrix} \mathrm{var}(\mathbf{x}) & \mathrm{cov}(\mathbf{x},\mathbf{y}) \\ \mathrm{cov}(\mathbf{x},\mathbf{y}) & \mathrm{var}(\mathbf{y}) \end{pmatrix}.$$

Note that, if $\bar{x}$ and $\bar{y}$ are zero, then $\mathbf{C}(\mathbf{d})$ is just equal to the matrix product $\mathbf{d}^\dagger\mathbf{d}$, where $\mathbf{d}^\dagger$ stands for the transposed of $\mathbf{d}$. The next theorem is stated in terms of the determinant and trace of the covariance matrix $\mathbf{C}(\mathbf{d})$:

Theorem 6. *For an arbitrary 2-column datamatrix* $\mathbf{d}$ *one has*

$$\rho_E(\mathbf{d}) = |\max_{\mathbf{q}} \rho_P(\mathbf{dq})| = \sqrt{1 - \frac{4\det[\mathbf{C}(\mathbf{d})]}{\big(\mathrm{trace}\,[\mathbf{C}(\mathbf{d})]\big)^2}}$$

where the maximum is taken over all rotations $\mathbf{q}$. *A rotation* $\mathbf{q}$ *is maximizing if and only if the columns in* $\mathbf{dq}$ *are of equal variance.*

Proof. To maximize the expression $|\,\rho_P(\mathbf{dq})|$ it will prove to be comfortable to present Pearson's coefficient for an arbitrary 2-column matrix $\mathbf{d}$ as

$$|\rho_P(\mathbf{d})| = \sqrt{1 - \frac{\det[\mathbf{C}(\mathbf{d})]}{\mathrm{var}(\mathbf{x})\mathrm{var}(\mathbf{y})}}$$

where $\det[\mathbf{C}(\mathbf{d})]$ stands for the determinant of the covariance matrix of $\mathbf{d}$. To start the reasoning, note that for two arbitrary real numbers a and b one always has

$$ab \leq \left(\frac{a+b}{2}\right)^2$$

with equality if and only if a and b are equal. Taking for a and b the numbers $\mathrm{var}(\mathbf{x})$ and $\mathrm{var}(\mathbf{y})$, one arrives at the following inequality for Pearson's coefficient

$$|\rho_P(\mathbf{d})| \leq \sqrt{1 - \frac{4\det[\mathbf{C}(\mathbf{d})]}{\big(\mathrm{var}(\mathbf{x}) + \mathrm{var}(\mathbf{y})\big)^2}} = \sqrt{1 - \frac{4\det[\mathbf{C}(\mathbf{d})]}{\big(\mathrm{trace}\,[\mathbf{C}(\mathbf{d})]\big)^2}} \qquad (*)$$

with equality if and only if the columns of $\mathbf{d}$ are of equal variance. Now note that for an arbitrary 2×2 matrix $\mathbf{q}$ one has

$$\mathbf{C}(\mathbf{dq}) = \mathbf{q}^{\dagger}\mathbf{C}(\mathbf{d})\mathbf{q}$$

where $\mathbf{q}^{\dagger}$, as before, stands for the transposed of $\mathbf{q}$. It follows from this that in the special case where $\mathbf{q}$ is a rotation the right side of $(*)$ remains unchanged when replacing $\mathbf{d}$ by $\mathbf{dq}$. As a consequence one has for an arbitrary datamatrix $\mathbf{d}$ and an arbitrary rotation $\mathbf{q}$ that

$$|\rho_P(\mathbf{dq})| \leq \sqrt{1 - \frac{4\det[\mathbf{C}(\mathbf{d})]}{\big(\mathrm{trace}\,[\mathbf{C}(\mathbf{d})]\big)^2}}$$

with equality if and only if the columns in $\mathbf{dq}$ are of equal variance. It is easy to see that to every $\mathbf{d}$ there is always some rotation $\mathbf{q}$ that brings the latter about. Namely, when applying a rotation over an angle of $90°$ the variances of the first and second column are exchanged. Hence, by continuous rotation from $0°$ to $90°$ the column of smallest variance eventually switches over into the one of largest variance. It follows that there must be some angle in between $0°$ and $90°$ for which the columns are of equal variance. This proves that the maximum value of $|\rho_P(\mathbf{dq})|$ is equal to the right side of the above. □

Another explicit expression for the Euclidean correlation coefficient is given by the next theorem:

Theorem 7. *For an arbitrary $n \times 2$ datamatrix one has*

$$\rho_E(\mathbf{d}) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalue of the covariance matrix of $\mathbf{d}$.

Proof. This is a direct consequence of the previous theorem. □

The theorem above allows for an interesting interpretation of the Euclidean correlation coefficient in orthogonal regression theory. Orthogonal regression distinguishes itself from ordinary regression in that the orthogonal residual of a point in the plane relative to a straight line is defined as the Euclidean distance of the point to the orthogonal projection

of the point on the line in question, that is, the distance of the point to the line. In ordinary regression one takes the distance to the vertical projection. Given a cloud of points in the plane and a straight line, one could square the orthogonal residuals of all points and add them up. Thus the so-called *residual sum of squares* or the *sum of squares of errors* of the cloud of points relative to a straight line can be computed. The straight line that minimizes the residual sum of squares relative to a fixed cloud of points, completely similar to ordinary regression, is called the *regression line* in orthogonal regression theory. Regression lines in ordinary and orthogonal regression usually differ. To see how a regression line in orthogonal regression can be computed, suppose a straight line with equation $ax + by = c$ is given, where a and b are normalized such as to have $a^2 + b^2 = 1$. Then the residual sum of squares of a set of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ relative to this line is equal to

$$\sum_i (ax_i + by_i - c)^2.$$

The above can be rewritten as

$$\sum_i \big(a(x_i - \bar{x}) + b(y_i - \bar{y}) + a\bar{x} + b\bar{y} - c\big)^2.$$

This, in turn, can be rewritten as

$$n\left((a \;\; b)\,\mathbf{C}\begin{pmatrix} a \\ b \end{pmatrix} + (a\bar{x} + b\bar{y} - c)^2\right)$$

where $\mathbf{C}$ stands for the covariance matrix of the cloud of points. From the above it can be read off that the regression line must always pass through the barycenter of the cloud of points. Given this, the coefficients a and b can be computed by optimizing the matrix product

$$(a \;\; b)\,\mathbf{C}\begin{pmatrix} a \\ b \end{pmatrix}$$

under the constraint $a^2 + b^2 = 1$. It is well-known from linear algebra (see [10]) that the maximum and minimum are respectively the largest and smallest eigenvalue of the matrix $\mathbf{C}$ and that they are realized by (normalized) eigenvectors of this symmetric matrix. It thus appears that the underlying mathematics in orthogonal regression is equivalent to the determination of the principal components of the datamatrix $\mathbf{d}$ that belongs to the cloud of points (see [7]).

Now denote the largest and smallest eigenvalue of $\mathbf{C}$ by $\lambda_{\max}$ and $\lambda_{\min}$. From the above it follows that the residual sum of squares of the regression line is equal to $n\lambda_{\min}$. The fact that the regression line passes through the barycenter implies that the sum of the residuals is zero. This, in turn, implies that the variance of the residuals is equal to $\lambda_{\min}$. For this residual variance one always has

$$\lambda_{\min} \leq \frac{\lambda_{\max} + \lambda_{\min}}{2} = \frac{\mathrm{var}(\mathbf{x}) + \mathrm{var}(\mathbf{y})}{2}$$

with equality if and only if $\lambda_{\max} = \lambda_{\min}$. Equality would present the worst case scenario as to an orthogonal regression fit. It thus seems natural to define, similarly to ordinary regression, the concept of the *fraction of unexplained variance* as the quotient of the left and the right side of the above.

Definition 8. The *fraction of explained variance* in orthogonal regression is understood to be the number

$$1 - \frac{2\lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

In this terminology the fraction of explained variance is precisely the value of the Euclidean correlation coefficient. Recall that in ordinary regression it is the square of Pearson's coefficient that presents the fraction of explained variance.

6 Comparing bivariate measures of linear association

Given a measure of linear association $\mathbf{d} \mapsto \gamma(\mathbf{d})$ there are many ways to derive other measures from it. One trivial way to do this is by transforming the measure by means of a strictly increasing continuous function $f : [0, 1] \to [0, 1]$ with $f(1) = 1$. Namely, if γ is a measure of linear association then so is the composition $f \circ \gamma$.

Definition 9. Two measures of linear association γ_1 and γ_2 will be called *equivalent* if there exists a strictly increasing continuous function $f : [0, 1] \to [0, 1]$ such that $\gamma_2 = f \circ \gamma_1$.

Measures are sometimes created (see [2], [5]) by so-called 'balloon rules'. The idea of such measures is, heuristically, that an ellipse is fitted to the scatterplot in such a way that it contains, say, some 95% of the points. Then the axes of the ellipse are compared in the form of a ratio. Actually, all these measures are more or less based on the ratio of $\lambda_{\min}$ and $\lambda_{\max}$. To be more precise, if $f : [0, 1] \to [0, 1]$ is a strictly decreasing continuous function with $f(0) = 1$ then a map of the form

$$\mathbf{d} \mapsto f\left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)$$

presents a geometric measure of linear association. A subclass of geometric measures emerges:

Definition 10. A measure of linear association that can be captured in the way sketched above will be called an *elliptic* measure.

Generally all elliptic measures are equivalent. By the results in the previous section the Euclidean correlation coefficient is a special example of an elliptic measure. It follows that a measure is elliptic if and only if it is equivalent to the Euclidean correlation coefficient. A non-trivial example of an elliptic measure is provided by the following:

Example 11. Given an arbitrary datamatrix $\mathbf{d}$ one could consider to determine the mean value $\bar{\rho}_P$, with respect to the Haar measure on the group of rotations, of the absolute value

of Pearson's coefficient over all rotations **dq**. This mean can easily be computed. To see this, suppose that the datamatrix **d** has columns of equal variance and that its Pearson coefficient is ρ. If $\mathbf{d}'$ is a rotation of **d** over an angle φ, then Pearson's coefficient of $\mathbf{d}'$ is related to ρ as:

$$\rho_P(\mathbf{d}') = \frac{\rho\,\cos(2\varphi)}{\sqrt{1-\rho^2\sin^2(2\varphi)}}.$$

The above can be derived by elementary trigonometric arguments. Consequently

$$\begin{aligned}\bar{\rho}_P(\mathbf{d}) &= \frac{1}{2\pi}\int_0^{2\pi}\frac{|\,\rho\,\cos(2\varphi)|}{\sqrt{1-\rho^2\sin^2(2\varphi)}}\,\mathrm{d}\varphi \\ &= \frac{2}{\pi}\int_0^{\pi/2}\frac{|\,\rho\,|\,\cos(\theta)}{\sqrt{1-\rho^2\sin^2(\theta)}}\,\mathrm{d}\theta = \frac{2\arcsin(|\,\rho\,|)}{\pi}.\end{aligned}$$

However, the variances of the two columns in **d** being equal, one has (applying Theorem 6) that $|\,\rho\,| = \rho_E(\mathbf{d})$. It follows that for such datamatrices one has

$$\bar{\rho}_P(\mathbf{d}) = \frac{2\arcsin\left[\rho_E(\mathbf{d})\right]}{\pi}.$$

Both sides of the equality above being invariant under rotations, the equality holds for arbitrary 2-column datamatrices. Thus the mean of Pearson's coefficient appears to be elliptic. □

Non-elliptic geometric measures do exist:

Example 12. Let **d** be an arbitrary $n \times 2$ datamatrix and ℓ an arbitrary straight line in $\mathbb{R}^2$ that passes through the origin. Project (orthogonally) the points in $\mathbb{R}^2$ that are defined by **d** on the line ℓ. Thus a subset of ℓ is created. Denote the diameter of this subset by $\delta(\mathbf{d}, \ell)$. Define $\delta_{\min}$ and $\delta_{\max}$ as

$$\delta_{\min} = \min_{\ell}\,\delta(\mathbf{d},\ell) \qquad \text{and} \qquad \delta_{\max} = \max_{\ell}\,\delta(\mathbf{d},\ell)$$

where the minimum and maximum is taken over all straight lines ℓ that pass through the origin. A measure γ can now be defined by setting

$$\gamma(\mathbf{d}) = 1 - \frac{\delta_{\min}}{\delta_{\max}}.$$

This γ is a geometric measure of linear association. To see that the measure γ is not elliptic, consider the two datamatrices $\mathbf{d}_1$ and $\mathbf{d}_2$, defined by

$$\mathbf{d}_1 = \tfrac{3}{2}\sqrt{2}\begin{pmatrix}1 & 1\\ 0 & 2\\ 1 & 3\end{pmatrix}, \qquad \mathbf{d}_2 = \sqrt{3}\begin{pmatrix}-1 & 0\\ 1 & 1\\ 0 & -1\end{pmatrix}.$$

The matrices above are such that their covariance matrices both have eigenvalues 1 and 3. In spite of this one has

$$\gamma(\mathbf{d}_1) = \frac{1}{3} \quad \text{and} \quad \gamma(\mathbf{d}_2) = \frac{1}{4}.$$

It follows that γ cannot possibly be elliptic. □

Given some geometric measure of linear association, there might be a wish to convert it into an algebraic measure. A way to bring this about is by standardization (see also [2]). For an arbitrary vector $\mathbf{x}$ in $\mathbb{R}^n$ its standardization is defined as

$$\text{std}(\mathbf{x}) = \frac{1}{\sqrt{\text{var}(\mathbf{x})}} \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}.$$

For a datamatrix $\mathbf{d}$ its standardization $\text{std}(\mathbf{d})$ is understood to be the matrix that arises when standardizing all its columns. For an arbitrary measure of linear association γ one can in this way create a map

$$\mathbf{d} \mapsto \gamma\,[\text{std}(\mathbf{d})]\,.$$

This map presents an algebraic measure of linear association.

Definition 13. The map above will be called the *standardization* of γ and will be denoted as $\text{std}(\gamma)$.

When taking the Euclidean correlation coefficient as a candidate one arrives at:

Theorem 14. *The standardization* $\text{std}(\rho_E)$ *of* ρ_E *is equal to* ρ_P.

Proof. If $\mathbf{d}$ is a datamatrix with columns of equal variance, then one has (by Theorem 6) that

$$\rho_P(\mathbf{d}) = \rho_E(\mathbf{d}).$$

For an arbitrary datamatrix $\mathbf{d}$ the matrix $\text{std}(\mathbf{d})$ has columns of equal variance, hence one generally has

$$\rho_P(\mathbf{d}) = \rho_P\big(\text{std}(\mathbf{d})\big) = \rho_E\big(\text{std}(\mathbf{d})\big).$$

It follows from this that $\text{std}(\rho_E)$ is equal to ρ_P. □

Theorem 15. *Two geometric measures of linear association are equivalent if and only if their standardizations are equivalent. They are equal if and only if their standardizations are equal.*

Proof. If two geometric measures γ_1 and γ_2 are equivalent then evidently $\text{std}(\gamma_1)$ and $\text{std}(\gamma_2)$ are so. To prove the converse, suppose that $\text{std}(\gamma_1)$ and $\text{std}(\gamma_2)$ are equivalent. Then there exists a continuous strictly increasing function $f : [0, 1] \to [0, 1]$ such that $\text{std}(\gamma_1) = f \circ \text{std}(\gamma_2)$. Using the fact that a measure of linear association is invariant under scalar multiplications, this actually implies that one has

$$\gamma_1(\mathbf{d}) = f\big(\gamma_2(\mathbf{d})\big)$$

whenever the columns of $\mathbf{d}$ are of equal variance. Now let $\mathbf{d}$ be an arbitrary datamatrix and let $\mathbf{q}$ be a rotation such that $\mathbf{dq}$ has columns of equal variance. Because both γ_1 and γ_2

are geometric, one may write

$$\gamma_1(\mathbf{d}) = \gamma_1(\mathbf{dq}) = f\big(\gamma_2(\mathbf{dq})\big) = f\big(\gamma_2(\mathbf{d})\big).$$

This proves that $\gamma_1 = f \circ \gamma_2$ and thus the equivalence of γ_1 and γ_2. The second statement in the theorem can be proved in exactly the same way. □

The next theorem contributes in explaining the prominent role of Pearson's coefficient in bivariate statistics.

Theorem 16. *A measure of linear association is equivalent to Pearson's coefficient if and only if it is the standardization of an elliptic measure.*

Proof. Just apply the two previous theorems. □

In the light of the 'balloon constructions' of measures, the theorem above explains why in the past many efforts to create new bivariate correlation coefficients ended up with measures equivalent to Pearson's coefficient (see [2], [5]). To have an example of an algebraic measure that is not equivalent to Pearson's coefficient, just standardize the non-elliptic measure in Example 12.

The idea to exploit group theory in the creation and classification of measures of linear association turns out to be a powerful tool. More details can be found in [8], where these techniques are discussed in detail when dealing with multivariate measures.

References

[1] Bourbaki, N.: *Algèbre*. Hermann, Paris 1965.

[2] Coleman, B.J.: A Coefficient of Linear Correlation Based on the Method of Least Squares and the Line of Best Fit. *The Annals of Mathematical Statistics* 3 (1932) 2, 79–85.

[3] Curtis, M.L.: *Matrix Groups*. Springer-Verlag, Berlin 1998.

[4] Gebelein, H.: Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *ZAMM, Z. Angew. Math. Mech.* 21 (1941) 6, 364–379.

[5] Loh, W.Y.: Does the Correlation Coefficient Really Measure the Degree of Clustering around a Line? *J. Educational Stat.* 12 (1987) 3, 235–239.

[6] Moore, D.S.; Notz, W.I.: *Statistics, concepts and controversies*. W.H. Freeman & Company, San Francisco 2006.

[7] Pestman, W.R.: *Mathematical Statistics*. 2nd ed., Walter de Gruyter Verlag, Berlin 2009.

[8] Pestman, W.R.: Creating and Classifying Measures of Linear Association by Optimization Techniques. *Math. Scand.* 106 (2010).

[9] Sénéchal, B.: *Groupes et géométries*. Hermann, Paris 1979.

[10] Voïévodine, V.: *Algèbre linéaire*., Editions Mir, Moscou 1976.

Wiebe R. Pestman
University Medical Center Utrecht
Julius Center (location Stratenum)
Universiteitsweg 100
3584 CG Utrecht, The Netherlands
e-mail: wpestman@umcutrecht.nl