Reminder of the Directional Nature of the Product-Moment Correlation Coefficient

Article · June 2022

DOI: 10.20935/AL5313

Jari Metsämuuronen
FINEEC Finland

ACADEMIA | Letters

Reminder of the Directional Nature of the Product–Moment Correlation Coefficient

Jari Metsämuuronen, Finnish Education Evaluation Centre, 2) Centre for Learning Analytics, University of Turku, Finland

Abstract

Product–moment correlation coefficient (PMC), "Pearson correlation", is usually taken as a symmetric measure of association because it produces equal estimate irrespective of how two variables in the analysis are declared. However, when the scales of the variables differ from each other, PMC is unambiguously a directional measure directed so that the variable with a wider scale (*X*) explains the order or response pattern in the variable with a narrower scale (*g*) and not in the opposite direction nor symmetrically. Hence, whenever the statistic *r*2 is calculated as is usually seen in the general scatterplots, this statistic is always a directional measure if the scales differ from each other; this should be kept in mind when interpreting *r*-squared statistics.

Keywords: product–moment correlation coefficient, coefficient eta, directional coefficient, eta squared, Goodman–Kruskal *G,* Somers *D*

Many faces of PMC

Of the coefficients of association between two variables in a general use, product moment correlation coefficient (PMC; Pearson, 1896 onwards) has the longest history and the widest applicability, and, maybe because of these, it is referred to in academic works far more times

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.

than the other coefficients taken together. Pearson himself valued the *polychoric* correlation coefficient generalized from tetrachoric correlation (Pearson 1900 onwards) as being one of his most important contributions to the theory of statistics (see Ekström, 2011). However, the fact is that PMC turned out to be the very tool for every researcher all over the world. PMC is firmly embedded in many routines such as procedures related to regression analysis, factor analysis and structural equation modelling, in estimating reliability, and in item analysis as one of the classical estimators of item discrimination power. Without exaggeration, we may say that PMC has been, and it still is, one of the main engines of many practical applications in the modern scientific inquiries and analysis settings.

Symmetric and directional measures of association

The coefficients of association can be divided into symmetric and directional ones although the directionality itself is often defined loosely. In the practical settings related to the estimation of correlation, the directional measures give (usually) two options for the association: either of the variables is independent and the other is dependent while the symmetric measures handle two variables as independent ones and they give, traditionally, only one estimate of the association. Some traditional directional measures of association are Goodman–Kruskal lambda and tau (Goodman & Kruskal, 1954), Somers delta (*D*, Somers, 1962), and coefficient eta (η, Pearson, 1903, 1905) and some with a symmetric nature are Pearson phi (Pearson, 1904) and Kendall tau-a and tau-b (Kendall, 1938, 1948). Notably, Metsämuuronen (2021) showed that also Goodman–Kruskal gamma (G), which also is considered traditionally as a symmetric measure because of producing only one estimate, is factually a directional measure. PMC is traditionally taken as a symmetric measure because it produces only one estimate for the association; the variable labels, e.g., *X* and *Y*, are arbitrary because the measure takes on the same value no matter how two variables are declared, that is, *PMCXY* = *PMCYX* (e.g., Walk & Rupp, 2010). However, it is to be seen that, factually, PMC is a directional measure if the number of categories of two variables differ from each other.

Directional nature of PMC

In theory, PMC is a measure of observed association for two continuous variables (see a typology in Olsson, Drasgow, & Dorans, 1982). However, its computational mechanic is used also in such measures as point biserial correlation (*RPB*) between a binary variable and a metric variable (with an ordinal, interval, or continuous scale), point polyserial correlation coefficient (*RPP*) between an ordinal variable with a narrower scale and a metric variable with a

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.

wider scale, as well as in rank-correlation coefficient between two ordinal variables famously simplified by Spearman (1904).

The hidden directional nature of PMC is easy to show with point biserial correlation. A simplified form for *RPB* is found in textbooks (e.g., Lord, Novick, & Birnbaum, 1968; Metsämuuronen, 2017), is shown to be identical with a specific direction of a truly directional coefficient eta (*X* dependent) and not with the other direction (*g* dependent) (Wherry & Taylor, 1949; Metsämuuronen, 2022). In the polytomous ordinal case, that is, with the point polyseral correlation, the absolute value of PMC is always lower than eta (*X* dependent) although so the value of PMC follows closely the direction of "*X* dependet" rather than "*g* dependent" (see the reasons in Metsämuuronen, 2022). This is illustrated in Figure 1 based on a dataset published at http://dx.doi.org/10.13140/RG.2.2.17594.72641.

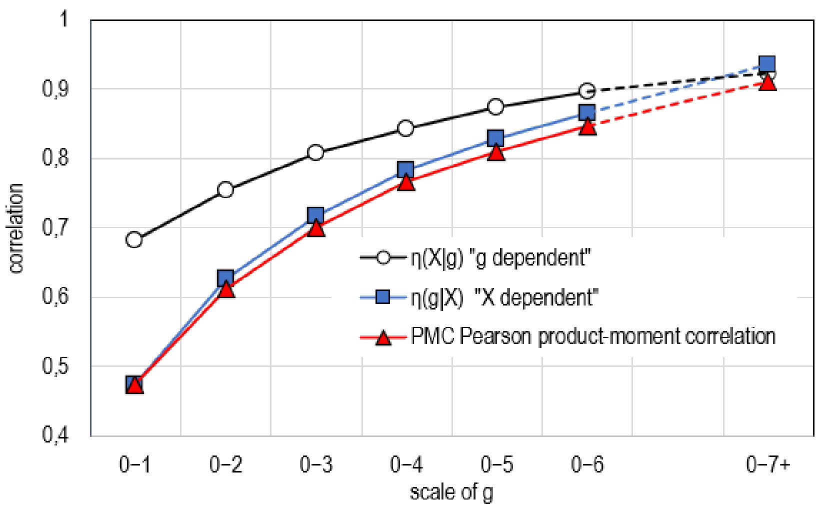

Figure 1. Connection of PMC and eta in real-life datasets (k = 14,880 estimates)

Relevance of the directionality in PMC in the practical settings

The bottom line is that because eta is a truly directional coeffcient, and PMC follows only *one* of the two directions of eta, *PMC is not a symmetric measure of association if the number of categories of variables differ from each other*. This far, the phenomenon is clear. The question is what the consequences of the directionality in PMC are. On the one hand, for many

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.

practical settings, the results have no effect. For example, if the variables are (essentially) continuous, or when the variables have an identical scale, directionality is not relevant issue. Also, if the number of categories in the scales are close each other and the sample size is high, the directionality may not have a practical relevance—the issue is merely in principle rather than in practice. For the other hand, when there is a radical asymmetry in the scales of two variables, PMC is always unambiguously a directional measure, and this may have notable practical consequences. Also, $R2$ seen in the standard scatterplots of the bivariate correlation or in the ordinary least square regression is not a symmetric but a directional statistic so that the *variable with a wider scale explains the order in the variable with a narrower scale* (see the discussion of the paradoxical meaning of the direction "*X* dependent" in Metsämuuronen, 2022).

In different types of settings related to the use of correlation coefficients, the interpretations of the phenomenon varies. Within the general settings of *correlational studies*, the explaining power by PMC between a variable with a narrower scale (g) and a wider scale (X), $r2$ indicates the extent to which X explains g. In the language of conditions, this can be expressed by using the phrase "g given X" (see discussion of the directions in Metsämuuronen, 2020, 2022).

Within the *settings related to GLM* with dichotomous independent variables, the direction of PMC equals with the traditional direction related to $\eta2$, and, in the polytomous ordinal case, it corresponds closely to the traditional direction of $\eta2$ directed the way we usually use $\eta2$ in the GLM settings. In these settings, the opposite naming ("*X* dependent") is traditionally used for the same direction as above. However, in these settings too, in the case of binary or ordinal settings, X explains the *order* of the responses in g and hence with a dichotomous and ordinal g, we could call this also "g given X", that is "g dependent".

Within the *settings related to measurement modelling* with items (g) and scores or measurement scales (X), the explaining power by PMC ($r2$) is always related to the direction when the score or measurement scale explains the response pattern in an item—not the other way around nor symmetrically; here also the expression related to conditions ("g given X") is relevant. This means that the traditional item–total correlation is a directional measure that indicates how well the score explains the response pattern in the item. This makes sense in the measurement modelling settings (e.g., Byrne, 2016; Metsämuuronen, 2017) and, hence, the directional nature of *RPB* and *RPP*, or item–total correlation, can be taken as a positive matter.

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.

References

Byrne, B. M. (2016). Structural equation modeling with AMOS. Basic concepts, applications, and programming. 3rd Edition. Routledge.

Ekström, J. (2011). The generalized definition of the polychoric correlation coefficient. UCLA, *Department of Statistic Papers*. https://escholarship.org/uc/item/583610fv

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764. http://dx.doi.org/10.1080/01621459.1954.10501231

Kendall, M. (1938). A new measure of rank correlation. Biometrika, 30(1/2), 81–93. http://dx.doi.org/10.2307/2332226

Kendall, M. G. (1948). Rank correlation methods (1st ed.). Charles Griffin and Co Ltd.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores.* Addison –Wesley Publishing Company.

Metsämuuronen, J. (2017). Essentials of research methods in human sciences. Volumes 1 3. SAGE Publications.

Metsämuuronen, J. (2020). Somers' *D* as an alternative for the item–test and item–rest correlation coefficients in the educational measurement settings. *International Journal of Educational Measurement, 6*(1), 207–221. https://doi.org/10.12973/ijem.6.1.207

Metsämuuronen, J. (2021). Directional nature of Goodman–Kruskal gamma and some consequences. Identity of Goodman–Kruskal gamma and Somers delta, and their connection to Jonckheere–Terpstra test statistic. *Behaviormetrika, 48*(2). http://dx.doi.org/10.1007/s41237-021-00138-8

Metsämuuronen, J. (2022). Artificial systematic attenuation in eta squared and some related consequences. Attenuation-corrected eta and eta squared, negative values of eta, and their relation to Pearson correlation. *Behaviormetrika.* https://link.springer.com/content/pdf/10.1007/s41237-022-00162-2.pdf

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika, 47*(3), 337–347. http://dx.doi.org/10.1007/BF02294164

Pearson, K (1896). VII. Mathematical contributions to the theory of evolution. III. Regres-

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.

sion, heredity and panmixia. Philosophical Transactions of the Royal Society of London, 187, 253–318. https://doi.org/10.1098/rsta.1896.0007

Pearson, K. (1900). I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 195*(262–273), 1–47. https://doi.org/10.1098/rsta.1900.0022

Pearson, K. (1903). I. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 200*(321–330), 1–66. https://doi.org/10.1098/rsta.1903.0001

Pearson, K. (1904). On the theory of contingency and its relation to association and normal correlation. Drapers' Company Research Memoirs. *Biometric Series I*, *XIII*. Soho Square, W.: Dulau and Co. http://archive.org/details/cu31924003064833

Pearson, K. (1905). On the general theory of skew correlation and non-linear regression. London. Dulau and Co. https://archive.org/details/ongeneraltheory00peargoog/page/n3

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27*(6), 799–811. http://dx.doi.org/10.2307/2090408

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101. http://dx.doi.org/10.2307/1422689

Walk, M. J. & Rupp, A. A. (2010). Pearson product-moment correlation coefficient (pp. 1022–1026). In N. J. Salkind, (Ed.), *Encyclopedia of research design*. SAGE Publications. http://dx.doi.org/10.4135/9781412961288.n309

Wherry, R.J., & Taylor, E.K. (1946). The relation of multiserial eta to other measures of correlation. *Psychometrika 11***,** 155–161. https://doi.org/10.1007/BF02289296

Corresponding Author: Jari Metsämuuronen, jari.metsamuuronen@gmail.com
Citation: Metsämuuronen, J. (2022). Reminder of the Directional Nature of the Product–Moment Correlation Coefficient. *Academia Letters*, Article 5313.