66-10

Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

"A Coefficient of Directional Correlation for Time Series Analyses"

by

ROBERT F. STRAHAN

MSDM
P95
qR311r

Report Number PR-66-10

THE LIBRARY

MSDM
P95
qR311r

University Archives

MSDM
P95
qR311r

A Coefficient of Directional Correlation for Time Series Analyses[1]

by

Robert F. Strahan[2]

December 30, 1966

PR-66-10

[1]Contributors to the preparation of this paper include R.D. Miller, C.K. Strahan, C.G. Katzenmeyer, and the major advisor of the writer, Dr. D.T. Lykken.

[2]Predoctoral research fellow, NIMH grant number 5 T1 MH-6964-04, Center for Personality Research, Department of Psychology, University of Minnesota.

Abstract

The coefficient of directional correlation indicates the degree to which the direction of change of two variables is the same from one sampling occasion to the next. This statistic is closely related to *Tau*, Kendall's coefficient of rank correlation, but in time series situations directional and rank correlation may be quite divergent. Application of the directional coefficient in the analysis of time series data is illustrated with an example from psychophysiology.

A Coefficient of Directional Correlation for Time Series Analyses

by

Robert F. Strahan

The tendency for two variables to change values in the same direction from one sampling occasion to the next can be termed "directional correlation." If increases in variable *x* are accompanied for the most part by increases in variable *y*, and decreases in *x* by decreases in *y*, then one can speak of some degree of positive directional correlation between variables *x* and *y*.

A quantitative expression for degree of directional correlation is readily derived. Suppose that values of *x* and *y* are recorded on *n* successive occasions. Comparison is made of the *x* and *y* values of each sampling occasion after the first with the *x* and *y* values of the immediately preceding occasion. A plus score is given for each instance in which the direction of *x* is the same as that of *y* (both variables increase in value or both decrease), a minus score for each instance in which the direction of *x* is opposite that of *y* (one variable increases, the other decreases), and zero or no score for each instance in which either or both variables remain unchanged. The coefficient of directional correlation ("directional" or "d" coefficient) is then defined as

$$d = \frac{p - m}{p + m}$$

where *p* is the number of pluses and *m* the number of minuses.[3] If there are no zero scores, $p + m = n - 1$, and *d* can equivalently be written as

$$d = \frac{p - m}{n - 1}$$

An illustration of the computation of _d_ is given in Table 1.

When the various intercorrelations among several variables are desired, a simple device makes possible greater efficiency in calculation. This device consists in first scoring pluses, minuses and zeroes for each variable separately, as that variable increases, decreases and stays the same in value, then calculating the various _d_'s on the basis of these preliminary scores. An illustration for three variables is given in Table 2. (This procedure also makes the computation of serial correlations a simple matter.)

It is easily seen that _d_ can range in value from - 1.0, indicating perfect negative directional correlation between the variables, to + 1.0, indicating perfect positive directional correlation. Values of _d_ about zero are evidence of little or no directional correlation in the sample.

The directional coefficient can be applied to various correlational situations. In the case of random sampling of independent observations from a given population, a close relation can be noted between _d_ and Kendall's coefficient of rank correlation, _Tau_. Both _d_ (Strahan, 1966) and _Tau_ (Kendall, 1955) are unbiased, consistent estimators of the population rank correlation coefficient, _Tau_'. Tau extracts more information from the sample and consequently is the more powerful statistic, while _d_ is considerably easier to calculate.[4] _d_ thus finds application when a short-cut estimate of population rank correlation is desired.

A more interesting application of _d_ is in the time series case. Here the sampling is not random, but temporal, and the observations often dependent rather than independent. In this situation _d_ may continue, along with _Tau_, to be an index of rank correlation. But directional correlation and rank correlation are conceptually distinct, and the distinction is most apparent in the case of time series. One may observe quite strong directional correlation between variables whose rank correlation is negligible or even opposite in sign, and vice versa. In such instances, both directional correlation and rank correlation are of interest, since each reflects a different aspect of the association between the variables.

An example of time series data with quite dissimilar directional and rank correlation is given in Figure 1. The plots of these electrodermal data of Maley (1967) show a strong association between the variables skin potential and skin conductance. That association, however, is clearly one of directional rather than rank correlation: d = 0.92, Tau = 0.10.

The source of divergence between directional and rank correlation is the former's insensitivity to relatively slow temporal changes in one or both variables. Although variability and sampling frequency are complicating factors, it is a fair generalization that, applied to pairs of variables changing over time, _d_ reflects phasic association to the exclusion of the tonic association present, while _Tau_ reflects tonic association to the neglect of phasic. Figures 2(a) and 2(b) are extreme examples constructed to illustrate this differential sensitivity of _d_ and _Tau_ to the two components of covariation. Figure 2(c) illustrates that directional and rank correlation may differ markedly even in the absence of any ready distinction between phasic and tonic components.

Footnotes

[3]This definition in effect omits from statistical consideration "tied" data--those portions of the data in which one or both variables exhibit no change in value. In certain applications the investigator may find it desirable not to exclude tied data from statistical representation. In such cases he should use the formula $d = \frac{p - m}{n - 1}$.

[4]The similarity of d and Tau is seen immediately in the case of untied data:

$$d = \frac{p - m}{n - 1} \qquad Tau = \frac{p - m}{\frac{n\,(n-1)}{2}}$$

where for both statistics p is the number of pluses and m the number of minuses scored for the sample of size n. Scoring for Tau, however, involves all possible x, y pairs in the sample -- $\frac{n(n-1)}{2}$ in number -- whereas scoring for d involves only a subset -- n - 1 in number.

References

Kendall, M.G. Rank correlation methods. (2nd ed.) London: C. Griffin, 1955.

Maley, M. J. A study of the relationship between three indicants of arousal: skin potential, skin conductance, and the two-flash threshold. Unpublished doctoral dissertation, University of Minnesota, 1967.

Strahan, R.F. The directional coefficient of correlation: a statistic related to Kendall's Tau. Unpublished manuscript, University of Minnesota, 1966.

Table 1. Illustration of the computation of d for a sample of six observations.

x	y	
6	4	
8	5	+
2	3	+
7	1	-
7	2	0
9	5	+

$$d = \frac{p - m}{p + m}$$

$$= \frac{3 - 1}{3 + 1}$$

$$= 0.50$$

Table 2 Illustration of a convenient scoring procedure when more than two variables are involved.

x		y		z	
6		4		3	
8	+	5	+	7	+
4	-	6	+	3	-
2	-	6	0	1	-
9	+	5	-	4	+
8	-	4	-	6	+

$$d_{xy} = \frac{2 - 2}{2 + 2} = 0.00$$

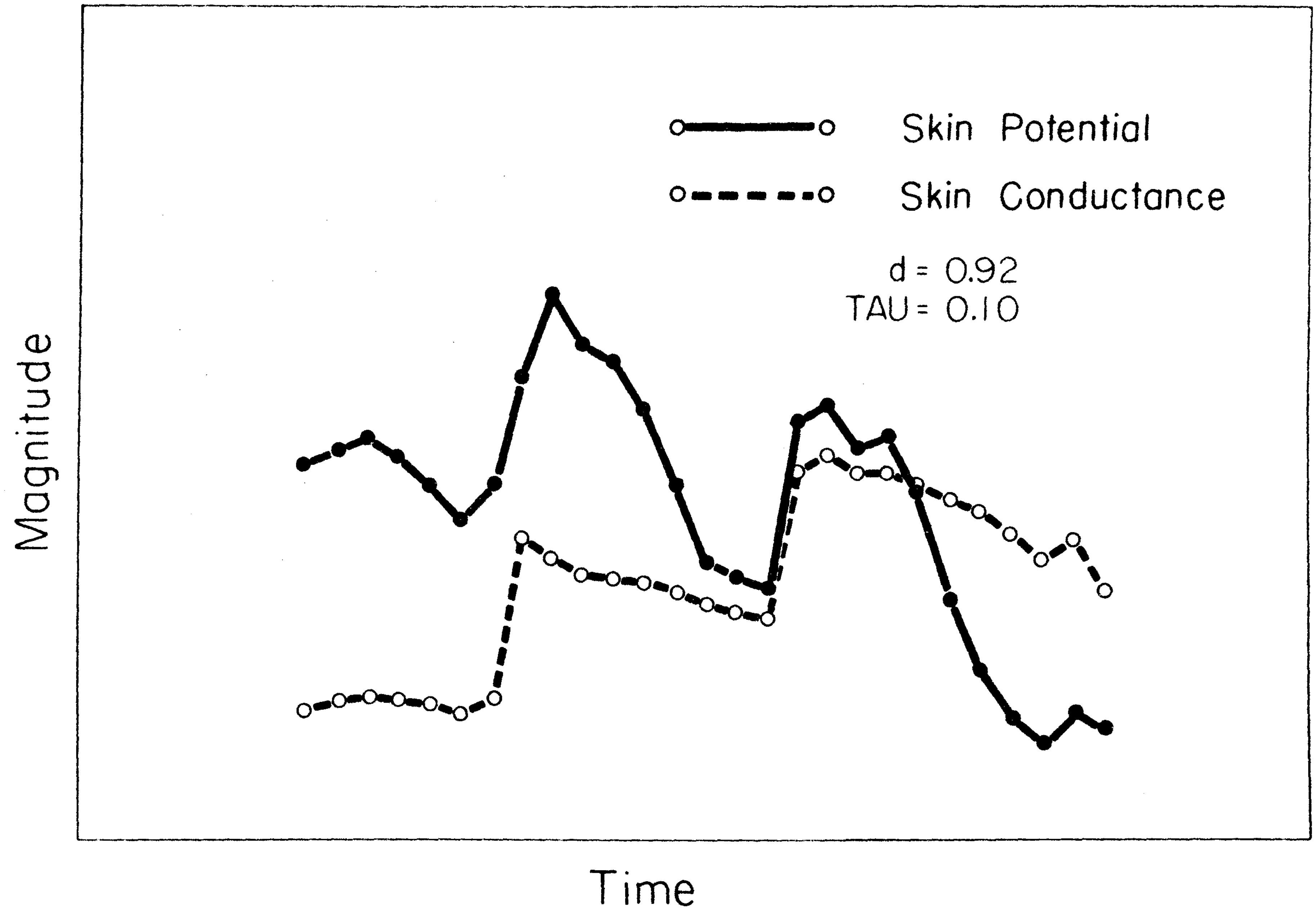
$$d_{xz} = \frac{4 - 1}{4 + 1} = 0.60$$

$$d_{yz} = \frac{1 - 3}{1 + 3} = - 0.50$$

Figure Legends

Figure 1. Example of divergent directional and rank correlation in psychophysiological time series data. The variables skin potential and skin conductance were recorded for a single subject over a 2-hour period.

Figure 2. Illustrative time series data showing greatly dissimilar directional and rank correlation.



Skin Potential
Skin Conductance
d = 0.92
TAU = 0.10
Magnitude
Time

FIGURE 1

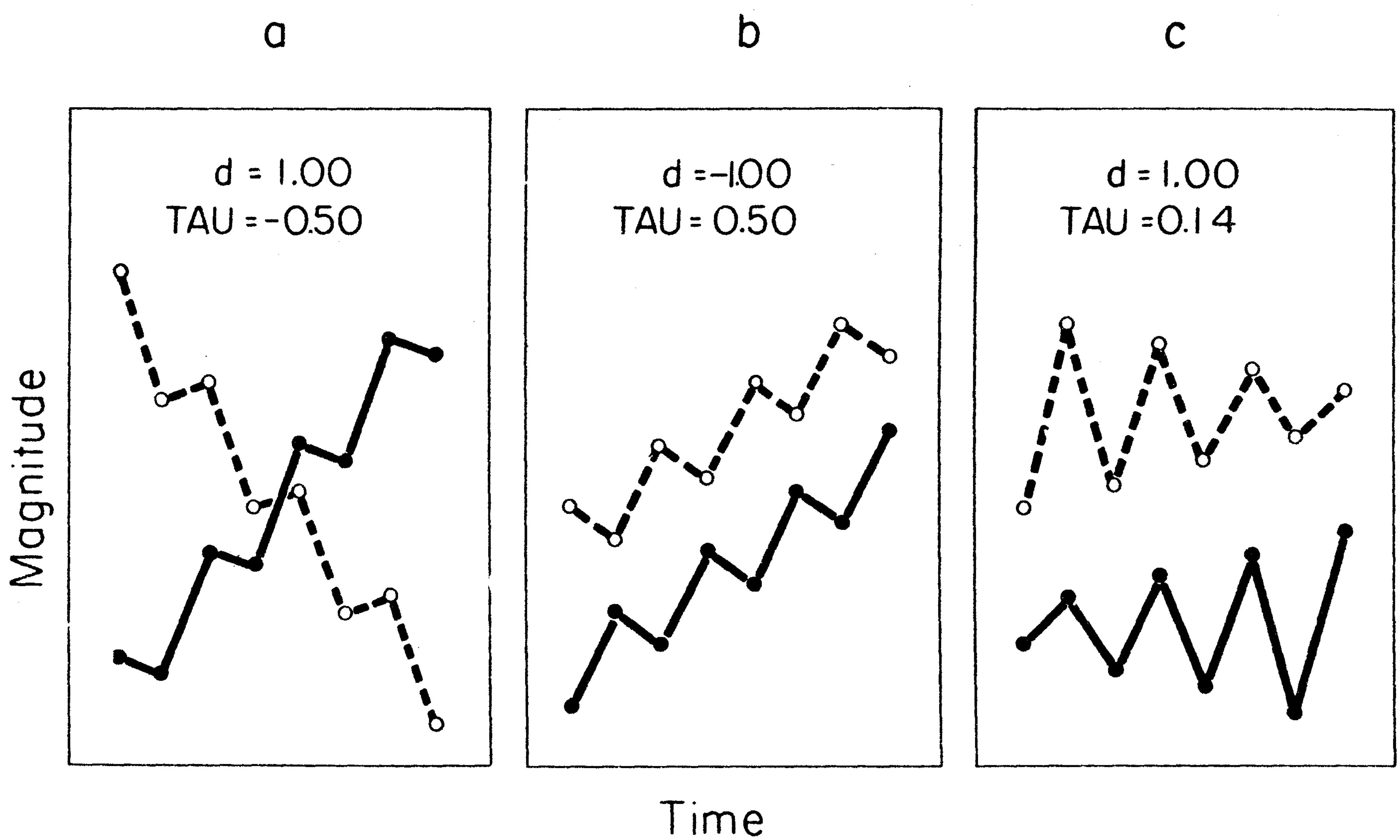


FIGURE 2