The spectre of ‘spurious’ correlations

D.A. Jackson and K.M. Somers

Department of Zoology, University of Toronto, Toronto, Ontario, Canada, M5S 1A1

Received March 2, 1990 / Accepted in revised form November 2, 1990

Summary. Ecologists often ‘standardize’ data through the use of ratios and indices. Such measures are employed generally to remove a ‘size effect’ induced by some relatively uninteresting variable. The implications of using the resultant data in correlation and regression analyses are poorly recognized. We show that ratios and indices often provide surprising and ‘spurious’ results due to their unusual properties. As a solution, we advocate the use of randomization tests to evaluate hypotheses confounded by ‘spurious’ correlations. In addition, we emphasize that identifying the appropriate null correlation is of utmost importance when statistically evaluating ratios, although this issue is frequently ignored.

Key words: Spurious correlation – Ratios – Statistical inference – Randomization tests – Data standardization

Recently, Prairie and Bird (1989, p. 287) stated that “the claim that the correlation between variables sharing a common term is spurious is a pervasive and unfortunate misconception within the ecological literature”. They suggest that the statistical community no longer regards the use of ratios in statistics as problematic and that the concept of a ‘spurious correlation’ is either dead or should be allowed to die. Although Prairie and Bird (1989) claim support from the statistical literature and a variety of other fields, they fail to distinguish the restrictive conditions under which ratio standardizations are appropriate from those where such standardizations result in ‘spurious correlations’ (sensu Pearson 1897). In fact, Aitchison (1986, p. xiii) states:

“Karl Pearson, in a now classic paper on spurious correlation, first pointed out dangers that may befall the analyst who attempts to interpret correlations between ratios whose numerators and denominators contain common parts ... History has proved him correct: over the succeeding years and indeed right up to the present day, there has been no other form of data analysis where more confusion has reigned and where more improper and inadequate statistical methods have been applied.”

Pearson (1897) was the first to identify the problems of using ratios in correlation analysis. He showed that two variables having no correlation between themselves would become correlated when divided by a third uncorrelated variable. The resultant correlation of these ratio variables or indices he termed a ‘spurious correlation.’ Specifically, he stated that a ‘spurious correlation’ was the “amount of correlation which would still exist between the indices, were the [variables] on which they depend distributed at random” (Pearson 1897, p. 490). Discussions of this ‘spurious relationship’ have appeared numerous times in the literature (e.g. Chayes 1971; Benson 1965; Atchley et al. 1976; Kenney 1982; Pendleton et al. 1983; Jackson et al. 1990). Prairie and Bird (1989, p. 286) state:

“Sokal and Rohlf (1981, p. 578) suggested that correlations between parts and wholes are ‘not really’ spurious, but are logical consequences of particular variable formulations. They suggested that there is no theoretical reason for avoiding such calculations, as long as the formulation is deliberate and well-considered.”

Although Sokal and Rohlf (1981, p. 578) suggest that the correlation of parts and wholes (e.g. total body weight and liver weight) is acceptable, Sokal and Rohlf also state that researchers should not be surprised to find “a sizeable correlation of a part with a whole” when the individual variables are uncorrelated and have a similar variance, a situation analogous to the condition which Pearson (1897) described. Sokal and Rohlf do not formally refer to the result as a ‘spurious correlation’, but rather as “logical consequences of particular variable formulation”. The contribution of “variable formulation” to the resultant correlation is widely recognized

Offprint requests to: D.A. Jackson
in the statistical literature, but is frequently ignored by those scientists who consider the concept of ‘spurious correlations’ to be misconstrued (e.g. Prairie and Bird 1989). The use of ratios in correlation analysis requires special statistical tests. This restriction is not identified by Prairie and Bird (1989). The necessity for the appropriate test and associated null hypothesis is demonstrated through examples below.

Why use ratios?

Ratios are prevalent throughout biology for two primary reasons (Phillips 1983). First, many ratios were developed in the early half of this century when computational difficulties and statistical theory limited more sophisticated analyses. Two or more variables could be incorporated into a single measure suitable for simple univariate or bivariate analyses. However, the widespread availability of computers now permits more powerful analyses (e.g. analysis of covariance; Packard and Boardman 1987, 1988). The second reason for using ratios is for data standardization. Frequently, a variable of interest may vary in magnitude with respect to a second variable of relatively little interest. For example, the fecundity of a fish generally increases with increasing body size. If one wishes to compare the fecundity of several fish differing in body size, a standardization is required to adjust for differences due to size alone (e.g. the gonadosomatic index; McQuinn 1989; Tanasichuk and Mackay 1989). Analogous measures are used in many fields (e.g. Harris 1970; Willis 1989). Unfortunately, such ratios correctly standardize for size only where the two initial variables (e.g. gonad weight and body weight) are collinear with an intercept of zero (Phillips 1983; Packard and Boardman 1987, 1988; Jackson et al. 1990). If either of these two conditions is not met, then the resultant relationship becomes ‘spurious’ (sensu Pearson 1897).

What is the null correlation?

Researchers generally regard the lack of association between two variables to result in a correlation of zero and test for association between two variables by constructing and testing a null hypothesis for X and Y of \( r_{XY} = 0 \). In this manner, one simply uses appropriate statistical tables to determine whether the observed correlation is sufficiently different from 0 to reject the null hypothesis. However, Pearson (1897) showed that the analysis of ratios may result in a null correlation differing substantially from zero. The sign and magnitude of the null correlation for ratios depends on the form of the ratio employed, the coefficients of variation of the raw variables, and the original correlation between the raw variables (e.g. see Chayes 1971, p. 11; Sokal and Rohlf 1981, p. 578; Kenney 1982). As a result, we cannot assume that a numerically large correlation implies significant association when using ratios (e.g. see Chayes 1971, p. 3).

To illustrate our point, we present the following example. Since the null hypothesis typically tested assumes no association between two variables, we will begin with a correlation of zero between X and Y (i.e. \( r_{XY} = 0 \)). We simulate data for X and Y such that the sample size (N) is 100, the mean (\( \mu \)) equals 100 and the standard deviation (\( \sigma \)) equals 30 for both variables. The sample correlation between X and Y is 0.09 and shows a bivariate normal distribution (Fig. 1a). Using X as our measure of size, we ‘standardize’ Y with respect to X producing the ratio Y/X. However, we now find that there is a strong non-linear correlation between Y/X and X induced simply by the ratio transformation (\( r = -0.66 \); Fig. 1c). As well, the ratio variable is no longer normally distributed. (We use Spearman rank correlation statistics throughout this paper due to the non-normal distributions and the non-linear relationships caused by ratios.) This ‘artificial’ correlation occurs because of the lack of independence between the ratio and its denominator (Meyer 1975, p. 171). Chayes (1971, p. 7) stresses that:

> “the null value against which an observed correlation is tested, will be precisely the correlation termed ‘spurious’ by Pearson – that is, the correlation that would be formed if the parent variables from which the ratios were formed were uncorrelated .... The null value, against which a sample correlation is to the tested, however, is not the zero characteristic of the untransformed variables but the correlation found after, and thus generated by, the transformation” (italics are Chayes).

Pearson (1897) calculated the expected correlation for two variables standardized by a third and a more general solution (Chayes 1971) is presented here:

\[
\left( \frac{CV; \sigma}{\mu} \right) \text{ for each variable is } \leq 0.15. \text{ If the } CV \text{ is larger, this approximation deteriorates and predicted null correlations can be as large as } r = 3.00 \text{ (see Atchley et al. 1976).}
\]

As an alternative to calculating the expected null correlation and subsequently assessing the associated probability of an observed correlation, we recommend using randomization tests (i.e. permutation tests; Sokal and Rohlf 1981, p. 787; Edgington 1987; Jackson et al. 1990). Normally, we evaluate the observed correlation (\( r_{obs} \)) using published statistical tables and the appropriate de-
Fig. 1. (a) Scatterplot of two independently simulated variables $X$ and $Y$ (rank correlation $r_{X,Y} = 0.09$). (b) Histogram of 2000 correlation coefficients from a randomization test of the variables $X$ and $Y$ from panel 1a. The observed correlation value is indicated by an arrow. (c) Scatterplot of $Y/X$ and $X$ showing a strong non-linear relationship ($r = -0.70$) induced by ratio formation. (d) Histogram of correlation coefficients from a randomization test of data from panel 1c. (e) Scatterplot of $Y/X$ and $Z/X$ showing a strong linear relationship ($r = 0.45$) induced by a common denominator in the ratios. (f) Histogram of correlation coefficients from a randomization test of data from panel 1e.

grees of freedom. Randomization tests (see Edgington 1987 for details) provide an alternative assessment of the probability without the restrictive assumptions associated with statistical tables. Quite simply, a randomization test allows one to determine whether $r_{obs}$ is greater than expected by chance (i.e. rejection of the null hypothesis of random association). Since biological data frequently violate the assumptions of classical statistics (e.g. random sampling, normal distribution), tabulated probabilities provide only an approximation of the correct probabilities that are generated by randomization tests (Edgington 1987).

The randomization test with normally distributed variables $X$ and $Y$ resulted in a distribution of expected correlation coefficients similar to a normal curve (Fig. 1b) and centered on zero. The corresponding ex-
pected distribution from correlations of \( Y/X \) and \( X \) is shifted to be centered at \(-0.70\) (Fig. 1d). Although there was no correlation between the parent variables \( X \) and \( Y \), the ratio shifted the expected distribution so that the usual null correlation of 0 no longer falls within the bounds of the expected distribution. That is, the expected value of \( \frac{r_Y}{X} \) given no association between \( Y \) and \( X \) is \(-0.70\), not 0. Thus, a test of \( r_Y X \) does not assess whether \( Y \) and \( X \) are randomly correlated (i.e. \( \frac{Y}{X} \) not associated with \( X \)).

If we consider the case where two variables \( Y \) and \( Z \) are 'standardized' by \( X \) (i.e. \( Y/X \) and \( Z/X ; r=0.45\), Fig. 1e), again we find that the usual null correlation of \( r=0 \) does not fall within the bounds of the expected distribution (Fig. 1f) even though \( X \), \( Y \) and \( Z \) were originally uncorrelated. Obviously, the traditional null hypothesis of a correlation of 0 is inappropriate if ratio variables are analyzed (see Jackson et al. 1990 for other published examples).

**When does a ratio remove the effect of size?**

There is a common perception that ratios are an acceptable standardization when there is a linear relationship between the numerator and denominator forming the ratio. For a ratio to remove the size effect, very restrictive conditions must be met (i.e. a linear relationship and intercept of zero). Such conditions must be evaluated with the parent variables (Packard and Boardman 1987, 1988). Prairie and Bird (1989, p. 288) state that the only time a correlation of ratios will provide a 'spurious' correlation is when a large measurement error is shared by both variables in the analysis. However, problems exist with ratios without addressing the difficulties resulting from measurement error (Long 1980, p. 64).

In addition, some researchers claim that ratios can be used to remove size effects provided the ratios are logarithmically transformed (e.g. Hills 1978). This transformation simply changes the relationship from \( Y/X \) versus \( X \) to that of \( \log Y - \log X \) versus \( \log X \) (i.e. the part-whole correlation of Prairie and Bird 1989). Although this transformation straightens the curvilinear relationships frequently observed with ratios, the expected or null correlation of \( \log(Y/X) \) versus \( \log X \) will differ from zero even though the parent variables (i.e. \( X \) and \( Y \)) are uncorrelated (see Jackson et al. 1990 for examples).

Prairie and Bird (1989, p. 287) erroneously assume the "correlation between such composite variables is always legitimate provided: 1) they satisfy the assumptions of correlation analysis, 2) the variables are meaningful, that is, they represent the concepts of interest and not just a component of them, and 3) the variables do not share a large measurement error term." Prairie and Bird (1989) fail to note that the null correlation changes when using ratios even though this condition has been long recognized (e.g. Pearson 1897; Chayes 1971; Mosimann 1962; Chayes and Kruskal 1966; Koch and Link 1971; Meyer 1975; Kenney 1982; Aitchison 1986; Jackson et al. 1990). Ratios correctly scale for size effects only in very special situations that are identified only with regression analysis! However, the same regression in the analysis of covariance is more powerful and less ambiguous than results derived from ratios (Green 1986; Packard and Boardman 1987, 1988).

Lastly, we reiterate that Prairie and Bird (1989) emphasize that the use of ratios is acceptable if the ratio is the variable of interest. Their argument hinges on the preconception that mathematical artifacts can be ignored if the null hypothesis explicitly addresses features of a standardized variable. They fail to acknowledge that the null relationship of no association is not a correlation of 0.0 when ratios are examined. In addition, they fail to recognize the difference between a variable that is measured in standardized units, and one that is constructed *a posteriori* by recombining several variables (see Jackson et al. 1990). In spite of whatever conceptual arguments that Prairie and Bird (1989) advance, the fact remains that the evaluation of hypotheses involving derived ratios requires the use of randomization-type tests. With or without a 'spurious' correlation, a randomization test will estimate the correct probability of observing the relationship in question by chance. Thus, the arguments of 'spurious' inference, legitimate hypotheses, and priority of concepts only serve to confuse rather than resolve the issue.

We feel that Prairie and Bird (1989) do not present a correct assessment of the implications of using ratios. As a result, they mislead researchers into believing that correlations of ratios do not present any difficulties. Although they suggest that 'spurious correlations' are a dead issue in statistics, we stress that, like the Phoenix, the concept has simply arisen under a different name (i.e. "the logical concept of particular variable transformations"; Sokal and Rohlf 1981, p. 578). Since ratios confound tests based on traditional statistical tables, we strongly recommend that researchers employ linear models such as analysis of covariance rather than resorting to ratios. However, if ratios are used, randomization tests should be employed, otherwise the analysis will result in incorrect statistical evaluations and subsequent erroneous conclusions.

**Acknowledgements.** We thank R.H. Green, K.A. Keenleyside, P. Legendre, and Y.T. Prairie for discussions leading to the development of this paper. This study was funded by a Natural Sciences and Engineering Research Council Scholarship to DAJ, an NSERC Operating Grant to Harold H. Harvey, an Ontario Renewable Resources Grant and an Ontario Ministry of the Environment Grant to HHH and DAJ.

**References**

Pearson K (1897) On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc R Soc London 60:489–498